PB84-122902

Guidelines for the Monitoring of Urban Runoff Quality

Urban Drainage and Flood Control District
Denver, CO

Prepared for

Municipal Environmental Research Lab.
Cincinnati, OH

Nov 83

U.S. Department of Commerce
National Technical Information Service
NTIS

EPA-600/2-83-124
November 1983

PB84-122902

GUIDELINES FOR THE MONITORING OF URBAN RUNOFF QUALITY

by

Michael B. Sonnen
The M. D. Sonnen Co., Inc.
Walnut Creek, California 94598

Grant No. R 806704-010

Project Officer

Douglas Ammon
Storm and Combined Sewer Section
Wastewater Research Division
Municipal Environmental Research Laboratory
Cincinnati, Ohio 45268

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

MUNICIPAL ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/2-83-124	2.	3. RECIPIENT'S ACCESSION NO. PB84 122902
4. TITLE AND SUBTITLE "Guidelines for the Monitoring of Urban Runoff Quality"		5. REPORT DATE November 1983
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Michael B. Sonnen		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Urban Drainage District 2480 West 26th Ave. Suite 156-B Denver, Colorado 80211		10. PROGRAM ELEMENT NO. CAZB1B
		11. CONTRACT/GRANT NO. R-806704
12. SPONSORING AGENCY NAME AND ADDRESS Municipal Environmental Research Laboratory - Cin, OH Office of Research & Development U.S. Environmental Protection Agency Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED Final Report 8/79-2/81
		14. SPONSORING AGENCY CODE EPA/600/14

15. SUPPLEMENTARY NOTES

Project Officer: Douglas C. Ammon - Telephone: (513)684-7871

16. ABSTRACT

This study addressed the specific field monitoring requirements for gathering data on the quality of storm runoff in urban areas. Particularly the study stressed the general need to state explicitly what the exact objective for field monitoring is to be in each case, because, as the study showed, the monitoring requirements and costs vary tremendously among possible objectives one might set for a monitoring program. The six objectives studied are: (1) scientific research on runoff quality change mechanisms, (2) problem identification monitoring, (3) preliminary design or alternative solution selection monitoring, (4) final design corroboration monitoring, (5) regulatory compliance monitoring, and (6) operational performance monitoring. The numbers of samples required in the example programs described for each of the six objectives ranged from 24 samples under the design objective to 54,000 samples under the scientific research objective. These numbers are to be appreciated to be illustrative only, but the difference in sampling intensity is remarkable. Comparable cost levels ranged from $11,200 for a one-year design monitoring objective to $9,340,000 for a five-year scientific research program covering three urban areas. A major conclusion of the report is that urban runoff quality behavior remains poorly understood, and it lags far behind the level of hydrologic and hydraulic sciences that have been brought to bear in the drainage and flood control areas.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Runoff, Combined Sewers, Storm Sewers, Water Quality, Cost Effectiveness, Mathematical Models, Contaminants, Methodology	Guidelines, Monitoring Sampling	13B

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 128
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

DISCLAIMER

Although the information described in this article has been funded wholly or in part by the United States Environmental Protection Agency through assistance agreement number R-806704-010 to The Urban Drainage & Flood Control District, it has not been subjected to the Agency's required peer and administrative review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

FOREWORD

The U.S. Environmental Protection Agency was created because of increasing public and governmental concern about the dangers of pollution to the health and welfare of the American people. Noxious air, foul water, and spoiled land are tragic testimony to the deterioration of our natural environment. The complexity of that environment and the interplay among its components require a concentrated and integrated attack on the problem.

Research and development are the necessary first steps in problem solution and they involve defining the problem, measuring its impact, and searching for solutions. The Municipal Environmental Research Laboratory develops new and improved technology and systems for the prevention, treatment, and management of wastewater and solid and hazardous waste pollutant discharges from municipal and community sources, for the preservation and treatment of public drinking water supplied, and to minimize the adverse economic, social, health, and aesthetic effects of pollution. This publication is one of the products of that research; a most vital communications link between the researcher and the user community.

The report that follows addresses the problem identification and impact measurement steps for a particular kind of wastewater, namely, urban storm runoff. The quality constituents in urban runoff and their interactions are only generally known and understood today. This work points out that many field monitoring purposes exist in the urban runoff field, and the monitoring requirements to satisfy each objective are determinable only from postulation of mechanisms being sought in the field and statistical tests for reliability of the answers obtained. The report gives guidance as to how many samples might be necessary and what the field and laboratory costs might be for satisfying any of the six objectives studied.

Francis T. Mayo
Director
Municipal Environmental Research
Laboratory

Preceding page blank

ABSTRACT

Requirements for field sampling of urban storm runoff quality, particularly the quality of runoff still on urban land and pavement surfaces, are related to statistical tests for reliability and to each of six possible objectives for monitoring that a sample program designer might set for his or her program. Costs associated with programs of field sampling, laboratory analyses, and data interpretation are also presented for each objective.

The six objectives studied are: 1) scientific research on runoff quality change mechanisms, 2) problem identification monitoring, 3) preliminary design or alternative solution selection monitoring, 4) final design corroboration monitoring, 5) regulatory compliance monitoring, and 6) operational performance monitoring.

The numbers of samples required in the example programs described for each of the six objectives ranged from 24 samples under the design objective to 54,000 samples under the scientific research objective. These numbers are to be appreciated to be illustrative only, but the difference in sampling intensity is remarkable. Comparable cost levels ranged from $11,158 for a one-year design monitoring objective to $9,336,343 for a five-year scientific reseach program covering three urban areas.

A major conclusion of the report is that urban runoff quality behavior remains poorly understood, and it lags far behind the level of hydrologic and hydraulic sciences that have been brought to bear in the drainage and flood control areas. Water quality modeling for urban runoff pollution control planning or design, therefore, should be viewed as rudimentary and possibly misleading, given the 1980 level of the state of that art.

This report was submitted in fulfillment of Grant No. R806704 by The Urban Drainage and Flood Control District and M. B. Sonnen Co. under the sponsorship of the U.S. Environmental Protection Agency. This report covers a period from January 1979 to March 1981.

CONTENTS

Page

Disclaimer ii

Foreword iii

Abstract iv

Figures vii

Tables viii

I An Overview 1

- The Project 1
- The Issues 1
- Study Scope and Objectives 4
- Study Organization 6
- This Volume 7

II Monitoring Objectives 9

- What About Storm Runoff? 9
- Objective I: Monitoring of Physical Processes -- Scientific Research 12
- Objective II: Problem Identification Monitoring 20
- Objective III: Alternative Solution Monitoring 25
- Objective IV: Design Monitoring 29
- Objective V: Regulatory Compliance Monitoring 32
- Objective VI: Operational Performance Monitoring 33
- References for Chapter II 35

III Previous Monitoring Programs 37

- Denver 37
- Houston 43
- Chicago 46
- San Francisco 50
- The Nationwide Urban Runoff Program 54
- Existing Monitoring Guidelines 55
- References for Chapter III 57

CONTENTS (Concluded)

Page

IV The Statistics of Monitoring 59

Sites and Samples 59
Quality Constituents 63
Assistance from Statistics 64
Other Statistical Tests 71
References for Chapter IV 79

V The Costs of Monitoring 80

Sampling in the Field 80
Laboratory Analysis 80
Data Interpretation 83
Inflation 84

VI Example Program Designs 87

Analyses and Samples 87
Comparative Costs of Complete Programs 87
Modified Plan for Objective I 87

VII Summary and Conclusions 97

Summary Discussion 97
Conclusions 98

Appendices 101
I Percentile Values for the Student's t Distribution 102

II Ninetieth Percentiles for the F Distribution 103

Ninety-Fifth Percentiles for the F Distribution 104

Ninety-Seven and One-Halfth Percentiles of the F Distribution 105

Ninety-Ninth Percentiles of the F Distribution 106

III Percentile Values for the Chi-Square Distribution 107

IV Statistical Guidance to Satisfaction of Six Urban Runoff Quality Monitoring Objectives 108

FIGURES

Number		Page
1	Recent Inflation of Lab Analyses Costs	85
2	Recent Inflation of Data Interpretation Costs	86
IV-1	Example Student's t Distributions	112
IV-2	Example Chi-Square Distributions	113

TABLES

Number		Page
1	Sampling Program Characteristics for Each Objective	60
2	Numbers of Sampling Sorties Required for Each Objective	61
3	Laboratory Analyses Assumed for Each Objective	62
4	Ratios of Degrees of Freedom to Chi-Square Which Appear in the Computation of Confidence Limits on Variance	78
5	Field Sampling Effort and Cost for Each Objective	81
6	Selected Laboratory Analyses Costs - 1980 Dollars	82
7	Summary of Field and Lab Programs for the Six Objectives	88
8	Laboratory Analyses Assumed and Lab Costs for Each Objective	89
9	Interpretive Analyses Costs and Effort for Each Objective	90
10	Total Annual Costs for Sampling and Interpretation for Each Objective	91
11	Total Period Costs for Monitoring to Satisfy Each Objective	92
12	Modified Objective I Program	94
13	Field, Laboratory, and Interpretation Effort for Modified Objective I	95
14	Annual Costs and Overall Present Worth for Modified Objective I Program	96

I. AN OVERVIEW

THE PROJECT

The study reported here addressed the specific field monitoring requirements for gathering data on the quality of storm runoff in urban areas. Particularly the study stressed the general need to state explicitly what the exact objective for field monitoring is to be in each case, because, as the study showed, the monitoring requirements and costs vary tremendously among possible objectives one might set for a monitoring program.

The study concentrated on "urban runoff" quality measurements, and "urban runoff" was taken to mean the rainwater collected on and flowing across land surfaces, streets, and gutters prior to its entry to either storm sewers or combined sewers carrying sanitary and storm wastewaters. Historically much sampling of runoff quality in urban areas has been practiced by sampling the contents or effluents from storm or combined sewers or drainage channels in which considerable integration of waters from various sources has already taken place. The majority of this work is directed to measurement of runoff quality in water still flowing on the land and street surfaces themselves, although flows and qualities in storm and combined sewers or at points of combined sewer overflows are discussed.

The study especially concentrated on statistical requirements and tests for sufficiency or significance of data sets as these statistical considerations bear on the degree of satisfaction of each possible monitoring objective. While insights to monitoring program design are to be gained through the use of statistical inferences and reduction of generic objectives to statistical hypotheses subject to test, there remains a considerable degree of art to specifying sample numbers, levels of confidence in the expected results, and the like. Particularly it is important that the investigator be able to anticipate the general behavior of the phenomena being sampled, or considerable sampling effort, time, and money could be wasted. These concepts are developed in detail throughout the report.

THE ISSUES

Urban storm runoff quality has been of topical interest nationwide since the middle 1960's. During the intervening years perceptions of alternative solutions to runoff's pollution potential have changed several times. In the early 1960's when combined sewer overflows were first reported to contain oxygen-demanding organics and suspended solids in concentrations rivaling those of untreated sanitary sewage, the national (and largely Federally-inspired) reaction was that separation of combined sewer systems to separate sanitary and storm sewers may become necessary.

In the middle 1960's to early 1970's, attention was focused, at first, on the apparent high cost and obvious inconvenience that such a major retrofitting would visit upon the nation's urban areas, particularly the older, larger, and most heavily developed cities. In the same period, pollution control emphasis shifted from treatment plant design criteria and effluent standards toward the setting of stream or receiving water criteria and enforcement of receiving water standards, which for good or ill were established at levels tantamount to the scientific criteria for water user tolerance.

From 1972 to the present, emphasis has been shifted again, back toward enforcement of pollution control legislation which focuses on source control of pollutant discharge and on levels of treatment to be applied to those sources which are deemed to be adequately safe and economically achievable universally.

During these philosophical swings in pollution control emphasis, urban runoff's relative importance as a potential source of pollution of the nation's waters has also been viewed in a variety of ways. In the 1960's it was believed (though never proved) to be a major pollutional source. In the early 1970's it was first modeled mathematically (though not exhaustively) so alternative measures of runoff control could ostensibly be evaluated. In the late 1970's, assessments of "needs" for control of storm and combined sewer discharges were made, and analyses of "nonpoint" sources of potentially polluting discharges were performed. At the opening of the 1980's it is now being generally thought that control of storm runoff pollution would cost so much money that it is not feasible to undertake major collection and treatment works, and storm (as opposed to combined) sewer discharges have been relegated to a low priority item. Combined sewer overflows, by contrast, which contain mixtures of storm runoff and sanitary sewage, continue to receive attention as potentially threatening sources of pollution.

It is interesting, if not surprising, that through all these changes in enforcement and control philosophy, there has never been a national program of significant size or duration to demonstrate conclusively and scientifically the degree to which urban runoff constituents present a serious threat or an environmental insult. While billions of dollars have been spent nationally and at state and local levels to build treatment plants for sewered municipal wastes, and 2.5 billion dollars have been spent on controls for combined sewer overflows, only small numbers of millions of dollars have been allocated to control of separate storm runoff pollution. (Of course, storm water control as a matter of drainage and flood protection has been a significant municipal budget item for years.)

While there are many reasons for the minor attention paid to separate runoff pollution, many of them probably good and sufficient, it can nonetheless be stated categorically that: Scientific information has not been collected and does not exist to demonstrate that the threats from storm runoff to downstream beneficial users is thousands of times smaller (millions/billions of dollars) than the threat from municipal sewage.

Moreover, it is curious to note, despite some recent Federal efforts of significant proportions, that local public works agencies have been left to fend for themselves as far as data collection, discharge criteria, control standards, and operational procedures related to storm water are concerned. The outlook for the future is that few changes in this regard are forthcoming.

The present state of affairs, then, is that storm runoff is handled at local levels largely through drainage and flood control infrastructures and mandates. Pollution control for this source of waste discharge is often vested nowhere in established communities, although newer areas or "urbanizing" regions often hand the problem to developers of "new" residential and commercial land to control through detention or retention basins whose designs are predicated almost exclusively on attenuation of peak flows and partial to complete entrapment of suspended soil and other particles. Maintenance requirements for such structures are perfunctory to nonexistent. Design review and construction inspection are virtually nonexistent as well, but those are largely topics for another study.

It is to be noted that Section 208 of the Water Pollution Control Act Amendments of 1972 (P.L. 92-500) mandated a nationwide attention to planning for the control of "nonpoint" sources of pollution, of which urban runoff is one. The 208 program of planning studies, funded with Federal money administered and granted by the U.S. Environmental Protection Agency (EPA), has run its course. No Federal money for implementation of the control plans identified in that process was made available by P.L. 92-500; none are anticipated; and local governments in general appear to be foregoing any unilateral implementation. A few exceptions exist wherein local programs require very little or no commitment of public funds.

There remains the overriding, inescapable truth -- often cited as compelling excuse for inaction -- that long-term records of scientific data do not exist to link urban runoff quality incontrovertibly to downstream damages to receiving water users, or to demonstrate that measures to abate pollution actually will cause improved water quality and improved utility or safety for subsequent beneficial users.

The study reported here was undertaken to address runoff quality in a general sense, as a potential source of pollution requiring control, and in a specific way to address the field monitoring or sampling necessary for a local government (or anyone) to undertake to answer any of a variety of questions which remain unanswered but which the future will inexorably raise.

STUDY SCOPE AND OBJECTIVES

There are virtually no things about urban runoff quality that are today fully understood, written down in available literature to be perused, and established as incontestable fact or scientific truth. Most people involved in this field of inquiry are aware of that. Accordingly, virtually all analyses of the topic, be they for purposes of research, engineering planning, design, operational control, or some combination of these, must include some period of field measurement. These measurements are made for any of a variety of reasons. They are made to characterize the magnitude and areal scope of an existing or potential urban runoff problem. They are made to obtain data necessary to calibrate mathematical models. They are made to aid decisions about which control methods would locally be the most cost effective. They are made to determine whether controls in place have been and continue to be effective. They are made to identify physical and chemical mechanisms that lead to chemical concentrations in runoff being what they are.

Eventually results from some of these studies are reported in a diverse body of technical literature. Because monitoring in the field is expensive and time consuming, all of us are hopeful that these reported results are building toward a consistent and transferable body of knowledge which can be used, in lieu of new monitoring in our town, to aid our control decisions. Unfortunately, this level of knowledge has not yet been attained. For one thing, the problem has too many facets, there are too many variables involved, for universal truths to be extracted from the relatively short period of monitoring thus far covered. Secondly, and perhaps more importantly, the studies reported to date have been singularly conceived and performed to the extent that results reported have not been transferable from place to place. In addition, no two investigators appear to measure the same things, or they measure them differently, or they measure some things and estimate others at various levels of accuracy or precision. The result is that the knowledge gained thus far largely remains a scattered, separated, unlinked set of local findings having little to do with runoff behavior or engineering solutions anywhere else.

Still further, not only are the data now available largely nontransferable, the ones collected are being stretched for all their expensive knowledge content, to the point of breaking down, to the point of misuse or even abuse. Mistakes are almost assuredly being made in planning and implementation of controls because meager data sets have been extrapolated to conditions which those data do not cover.

Study Scope

This study has addressed field sampling of urban runoff quality. It was intended to clarify for those contemplating a field survey what types and numbers of samples would be required, and what those measurements would cost, to assure reliable answers to any of a variety of questions one might pose about runoff quality. And it intended to make clear that the question asked, the objective for going to the field at all, will dictate fairly directly what the size and scope of the field measurement program must be. As a corollary, it purported to show that data collected for one purpose may be wholly inadequate to satisfy another purpose. Lastly, by discussing monitoring protocols for a variety of purposes in one place, it was hoped that the illustrative methodology developed would lead to a consistency of sampling protocols and reporting of results so that transferability of data gathered in diverse places would become possible.

Study Objectives

Specific objectives for the project were as follows:

1) List and Develop Monitoring Objectives. (Six possible objectives for going to the field to measure characteristics of urban runoff were to be chosen. These were:

1. Scientific research.
2. Problem identification.
3. Alternative solution monitoring.
4. Support of final design.
5. Regulatory compliance monitoring.
6. Operational performance monitoring.)

2) Review Current and Recent Monitoring Program Objectives and Successes. (Past monitoring programs reported in the literature were to be reviewed and the results and conclusions were to be compared with the programs' stated, original objectives.)

3) Review Literature on Monitoring Program Design. (Past works on monitoring of urban runoff were to be reviewed to indicate the state-of-the-art in monitoring protocol design.)

4) Review and Develop Monitoring Strategies for Receiving Waters and "Best Management Practices". (Special purpose monitoring for characterizing impacts of urban runoff on receiving waters and for planning or monitoring performance of certain runoff pollution control measures were to be reviewed. Guidance was to be given where weaknesses were apparent.)

5) Develop Monitoring Protocols for Satisfying Objectives. (Statistical inferences of reliable numbers of sampling sites and sample numbers were to be developed for each of the six monitoring objectives selected.)

6) Design and Produce a Guidelines Report. (A guidebook describing all the foregoing material was to be developed, the chief purpose of which was to enlighten those contemplating a field measurement program of urban runoff quality regarding the necessary or indicated minimum set of measurements they face to satisfy the objective they have set for themselves. This volume represents that guidebook, and the reader is left to judge the success attained in meeting that objective.)

STUDY ORGANIZATION

The work reported was performed under a U.S. Environmental Protection Agency grant (No. R806704010) offered to the Urban Drainage and Flood Control District, a regional body, in Denver, Colorado. Mr. L. Scott Tucker, the District's Executive Director, largely inspired the study's purposes, made inputs to the original proposal, directed the budgetary administration of the work, and reviewed the results. Mr. Ben A. Urbonas, the District's Chief Planning Officer, was responsible for day-to-day review of the project, coordination with EPA, and writing a summary of Denver's past experiences with runoff quality monitoring for this report. The majority of the work was conducted by Michael B. Sonnen, President, The M.B. Sonnen Co., Inc., of Walnut Creek, California, through a subcontracting agreement (No. 79-8.4) with the District.

Dr. Sonnen's organization subcontracted a number of tasks. The subcontractors included:

Camp Dresser & McKee Inc. (CDM)
Environmental Research Group, Inc. (ERG)
RAMLIT Associates
E.D. Driscoll Associates
Murray B. McPherson, Urban Water Management Consultant.

The assistance of Larry A. Roesner and Stergios Dendrou of CDM, Doug Robbins and Bill Prater of ERG, Yoram Litwin of RAMILT, Eugene Driscoll, and Murray McPherson is acknowledged. The contributions of each of these gentlemen were sizeable and significant.

For EPA, the study was directed by the Storm and Combined Sewer Section of EPA in Edison, New Jersey, and Cincinnati, Ohio. Mr. Douglas Ammon was the Project Officer, and he provided EPA's detailed attention to the study throughout its duration and made numerous suggestions for improvements of its outputs. Additionally Mr. Richard Field, Chief of the Section, provided review and conceptual advice. Mr. Anthony Tafuri was the original Project Officer, prior to his transfer to a Regional Office of EPA; and he reviewed the original proposal, made suggestions for improvements in scope and direction, and got the study started.

THIS VOLUME

This booklet has several purposes, and it can be read and used in several ways. Primarily it is intended as a guidebook, largely conceptual and philosophical, for study by those faced with a first or second sortie to the field to learn something of value about urban runoff quality.

This book is also a first of its kind. It represents a first look at measurements of runoff quality that could be taken for a variety of purposes. The "theoretical," mathematically stated precepts, therefore, may prove upon trial to be weak, incomplete, or even wrong. The work reported here will surely be improved and expanded as experience grows. It is believed, however, that the theoretical-looking expressions will not lead to under-measurement and the misjudgments that could cause. Errors, if there are any, should be on the side of over-measurement, but the writer judges that this is the correct side on which to err, at least for the immediate future. Still, the monitoring protocols developed here should be viewed as guidelines, not scientific truths, which they are not; and areas and degrees of flexibility and remaining judgments to be made should be apparent. Particularly, the reader might note the utility of statistical methods for aiding sample design, but he must also note the limited ability to total disutility of statistics for establishing scientific cause-and-effect relationships from the data gathered.

Lastly this book unabashedly reflects a frustration with current knowledge about runoff quality experienced by several District staff members, Dr. Sonnen, and a few of his subcontractors.

But it can be read merely to learn those feelings of a few individuals who have struggled for a number of years only partially successfully with this most intractable topic. This is just another way of repeating that the book does not purport to represent irrefutable logic or scientific truth. Hopefully it will serve as guidance, background, and food for thought.

Chapter II develops the list of six possible monitoring objectives, which are studied repeatedly in depth throughout the remainder of the book.

Chapter III describes previous monitoring experiences, including a recent one here in Denver, and it draws out the degree of success or failure of each of those programs to achieve its stated objectives. Where the reasons for failure can be deduced, they are.

Chapter IV describes the statistics of sample design and statistical tests that can be applied to a set of data to determine its variability, characteristics, and reliable information content.

Chapter V explores the costs of monitoring, which often are the limiting factors; but by being so limiting, they often lead to too little information being gathered to satisfy the fundamental objective.

Chapter VI puts the objectives of monitoring, statistical information, and cost data together to show the sample designs that might be used to satisfy each objective. The results show that some objectives take much longer, many more samples, and a great deal more money to satisfy than others.

Chapter VII presents our summary and conclusions. The appendices report detailed information about statistical methods and inferences.

II. MONITORING OBJECTIVES

WHAT ABOUT STORM RUNOFF?

Is urban storm water quality causing a problem, or not? For many years we have thought it was. Combined sewers were found to contain concentrations of biochemical oxygen demand (BOD) and suspended solids as high or higher than were found more often but in much smaller flows of untreated municipal sewage. Storm sewer flows of urban runoff have been reported to be laden with suspended solids and to contain moderate BOD loads, as well as the nutrients, nitrogen and phosphorus, and numerous heavy metals and other elements and organic compounds at nearly trace, but possibly significant, levels. So, "problems" have been perceived in some receiving waters from combined sewer overflows and separate stormwater discharges.

But are these "problems" requiring solution? If they are, why are these flows not treated just as sanitary flows are? Conventional thought has been that combined sewer overflows are a serious problem. They are believed to be serious because combined sewage is indeed human waste products and all its accompanying microorganisms and related disease-producing factors combined with less disgusting or threatening elements of sediment and nutrients in storm runoff. But the high flows during storm periods carry the untreated combined sewage directly into receiving waters and past the conventional treatment plants which have been sized to treat only the much smaller, average flows of sewage generated in dry periods by residential, commercial, and industrial sources alone. The obvious reason for not providing treatment for all the possible flows that occur, including peak storm flows, is the much higher cost of the larger treatment plants and/or storage capacity that would be required. Such expenditures are considered not only prohibitively large; but they are perceived to be largely wasted, since high flows occur only very infrequently, and hence much of the provided capacity would sit idle most of the time. So in relation to the cost to fix combined sewer overflow problems, for many years these "problems" did not seem large enough to justify the expense. In terms of separate storm runoff impacts on receiving waters, there has been a general finding of higher microorganism counts and depressed oxygen levels in receiving waters following storms, but these have not been judged to be compelling.

Because impacts persisted into the 1970's, grass roots concerns for environmental management eventually led to passage of the Federal Water Pollution Control Act Amendments of 1972 (P.L. 92-500), which for the first time mandated pollution control planning attention be given to "nonpoint" sources of pollution of which urban storm runoff was one. "Nonpoint", was coined to contrast with "point" sources of pollution which were the conventionally known discharges to receiving waters through discrete outfalls or pipes that occurred at specific "points." The term "nonpoint" was used to designate urban, rural, and agricultural runoff, which was more likely to enter receiving waters in sheets of flow along the edges of receiving waters where point, discrete controls were not feasible. Roadside ditches or drainage channels which discharge to larger receiving waters at distinct places were largely overlooked and fell "into the cracks." Jurisdiction and responsibility for these discharges remain clouded.

Denver, and specifically the Urban Drainage & Flood Control District -- a regional body encompassing more than the City of Denver, is particularly interested in this state of affairs, because it is an urban drainage agency with responsibility for managing the storm runoff through its urban and urbanizing region. Water quality concerns are not among its charges. But the physical works of this agency channel tremendous quantities of urban storm runoff through its member-agency communities and into the receiving waters of one of the most environmentally concerned states, Colorado.

The District is behooved, therefore, to take an active interest in storm runoff quality measurement and determination of the pollutional impacts, if any, that runoff may have. The District has participated in one study of runoff quality already in the Denver region, which in many respects proved inconclusive. (More is reported about this work in subsequent chapters.) Another study is underway now in the area, and the District is again participating, but only as a member of an advisory committee. The District involved itself in the work reported here, specifically because it perceived that measurements of "pollution" in urban runoff to date, at least in the Denver area, have not been related well enough to the questions the local governments in the Denver area may one day have to answer. Specifically, is urban runoff causing a pollution problem, if so, is that problem significantly large to justify remedial action, if so, what are the more cost-effective alternative solutions, how should the chosen solution be designed, and how should the chosen solution be monitored and operated for continued satisfactory performance?

At a 1979 conference on "Urban Stormwater and Combined Sewer Overflow: Impact on Receiving Water Bodies," sponsored by EPA, it was reported by several EPA personnel that 1) a national assessment of needs for storm and combined sewer pollution control showed that combined sewer overflow abatement alone would cost 26 billion dollars, 2) EPA and apparently the cities of the nation could not raise that amount of money for solving these problems, and hence they may not be as compelling as was once thought, and 3) since 1977 there has been a 5-year moratorium placed by Congress on the funding of construction grants by EPA for the building of separate storm runoff treatment systems. Moreover, it was indicated that for receipt of grants to alleviate combined sewer problems, applicants are required to show the cost-effectiveness of controls anticipated. Benefits to beneficial users of receiving waters in terms of water quality enhancements have to be shown over a 20-year period. In addition, applicants will be required to show the seasonality of benefits, the conditions with and without combined sewer overflow (CSO) facilities in place, and the percentage of time that violations will occur. While this sounds like a reasonable set of justifications and problem definitions, the monitoring required will be significantly large, and certainly greater in scope, depth, and cost than any current or previous monitoring.

In a 1980 prognostication about the future of the Clean Water Act [P.L. 92-500 and its 1977 amendments], Mauzy (1980)* has stated:

> Urban stormwater will remain as a potential problem; however much more has to be known about the cause-effect relationships and reasonable approaches to reduce pollutant loading to streams. Stormwater will be accepted as a less significant problem.
>
> Quite the opposite is true for combined sewer overflow, which will be recognized as a major pollution source. In fact, combined sewer overflow impacts will become the major constraint to achieving potential beneficial use of streams...

If much more needs to be known about storm runoff, how can the perception of its importance already be waning? We have yet to determine its importance for the first time. It may contain substances of greater threat to man than treated sanitary wastes. [It may not too; we don't know.] And it seems impossible to state "in fact" that combined sewer overflow will become either a major or a minor constraint, or none at all, on beneficial uses of streams (or lakes or estuaries) until and unless case-specific measurements are made to prove or dismiss the point.

In any event, it is clear enough that urban storm runoff is exposed to numerous contaminants, whether in a storm sewer or a combined sewer or still on the ground surface. It is not pure water. More measurements in more places and throughout the subregions of those places are necessary to learn just how urban storm runoff becomes contaminated, to what degree, and to what extent it poses a threat to man and the rest of the environment.

The whole point of this work, as will soon become evident, is that the type, number, and location of these measurements must be varied to match the specific objective for monitoring which is established in the first place.

---

*References are given at the end of each chapter in which they appear.

OBJECTIVE I: MONITORING OF PHYSICAL PROCESSES -- SCIENTIFIC RESEARCH

By "monitoring" we shall mean for the balance of this report: The collection of field samples, laboratory analyses of the samples' contents, and sufficient interpretation of the results to glean from the samples some useful information about urban runoff quality. No intent is implied to separate the "monitoring" referenced here from "surveillance" which others choose to distinguish from monitoring. The concern here is that scientific information on the processes occurring is not being collected thoroughly enough.

The concern for the bulk of this report is: Given a rainy day in any city and an inquisitive citizen wondering, officially or otherwise, "What's in that stuff?" how many samples would he have to collect and analyze? It is axiomatic that the answer depends on specifically why he cares about the things in runoff. If he wants to know how those things got there and in the quantities that exist there, he will have to sample more than if he wants to know how to get them out. He will have to sample to yet a different extent to determine what they will do to receiving waters.

In this section, it is assumed that the objective is to determine the physical and chemical processes, the environmental mechanisms, that cause concentrations of various ingredients in urban runoff to be what they are. There are several reasons for wanting to know such things. First, there is the fundamental research objective of merely seeking new knowledge, either for its own sake or to lead to conclusions about feasible control options or places in the natural cycle that can be interrupted or intercepted to effect some removal or abatement. Secondly, one might wish to construct a mathematical model of these processes, so that numerous alternative management schemes could be analyzed without having to construct each one and then to monitor each one in the field before choosing the most advantageous one. There are other reasons as well for pursuit of fundamental knowledge, but we shall limit our attention to these two.

Fundamental Runoff Processes

Much of the time in any urban area there is no rainfall or runoff at all. Storm drains lie darkly empty underground; and combined sewers or separate sanitary sewers carry their nominal, comfortable flows of domestic household wastes and industrial discharges through a predictable 24-hour pattern of flows. From 6:00 A.M. to noon, sewer flows rise as the city stirs and starts its chores. From noon to 6:00 P.M., flows fall off slowly as lunch time is passed, commerce and industry shut down, and dinner at home is prepared and served. From 6:00 P.M. to midnight, flows continue to wane as bedtimes are approached and the city largely calls it quits. From midnight to dawn, sanitary flows reach their minimum, solids begin to settle in pipes because turbulence of flow is so diminished, and suspended and organic loads reaching downstream treatment plants are at one-fortieth of their average hourly values. Only the new day's flushings and bathings will scour the sewers and bring flows and loads back up to their mean levels and beyond.

Outdoors on the ground and paved surfaces, on days without rain and runoff, there are stirrings and changes nonetheless. The wind is at work, moving particles of dirt and garbage that it *can* move; and bacteria and chemicals are at work making everything not only one day older but one day more reduced, more rotten, or simply more changed. On many surfaces of pavement, grass, brick, or dirt, the chemical leavings from the last storm, or the weathered bits of those same media themselves, wait patiently insoluble and trapped, stuck, until the next washing by a street sweeper or the coming of dissolving rain. The day and night temperatures stimulate and slow the reactions and microbial activity inexorably occurring.

And then, day or night, busy weekday or quiet weekend, it begins to sprinkle. The first drops of rain plink down, and a few spots of earth or roof or street or lawn begin to wet. What happens in the next few minutes is critically dependent on all that occurred and did not occur since it rained the last time. If a lawn has been fertilized and sprayed with fungicide or insecticide, these materials will begin to dissolve locally in drops of rainwater clinging to grass stems. If the street has been heavily travelled, accumulated metals from worn tires and exhaust fumes will begin to react with substances in the nearly, but not quite, pure rainwater; and they will dissolve in the water as ions to remain in that state until they are subsequently precipitated again as solids.

If rain continues to fall and to fall more rapidly, large areas of urban surface will become uniformly wet. Soil and pavement, shingles and tarpaper will absorb the water they can. Water will begin to flow down grass stems, shrubs, and trees to wet the ground beneath. If rain persists on pervious soil, grass-covered or not, soon water will begin to percolate to fill the porous material. Excess water the soil cannot accept will pile up until its mass can sense the downhill direction; and rivulets and sheets will begin a slow trickle along the ground surface toward a lower place. The same will occur without downward penetration, on paved or solid surfaces. Depressions nearby will trap the first dribbles, until they too are full, and general runoff will begin.

As flows of runoff build, very small soil particles and chemical precipitates will be loosed and buoyed by the running water and carried along. Depending on pH, temperature, and other controlling attributes of the water, some of these will also dissolve; but many will only partially dissolve and will partially be swept along as overwhelmed solids, unable to remain at rest. Five minutes into a storm, the rainwater could easily have changed from nearly pure water containing nitrogen, a little sodium, a trace of chloride, and undetectable organics scoured from the air or even sublimed during evaporation from a distant sea into a contaminated rivulet of runoff containing copper and zinc from auto exhausts, nitrate and phosphorus from fertilizers, arsenic and lead from paints and other decaying surfaces, and complex inorganic compounds from clay soils or exotic man-made poisons from pesticides.

With astonishing speed, before the rain storm has likely reached its peak intensity, runoff at nearby gutters on rooftops or at street curbs can be occurring. Particles stacked there in random bundles by the wind will be wet and will begin to yield to the water its more soluble components, which will leave immediately with the water that is still seeking a lower place. Solid bits of straw or earth or chemical precipitate will yield as well to the running water's growing force. These will become "suspended" solids whose mass cannot resist the turbulence of flow, and they will travel with the runoff until its force of movement diminishes sufficiently to lay them down again or until they are entrapped on some immovable object or mesh-like material on which they impinge.

Presently runoff in the city will plunge down a grated hole and find itself in an underground sewer. At the point of plunge, its quality may be slightly to considerably different from that of the rain still falling, like which it very recently was. By now it could contain tens of milligrams of numerous soluble anions and cations, nutrients and metals, organic compounds, oils and greases, as well as clay and bits of other simple dirts, hair and straw, and countless other kinds of urban jetsam.

If it has not begun to happen already, back up on the roof or lawn, then surely just before its plunge into a sewer at the end of a man-made block, a variety of physical-chemical processes can occur. A compound or ion from two inches or ten yards away which lay idle and unchanged for days may encounter now another in the mixture of runoff with which it has an unusual affinity. They may switch anionic partners and form new compounds. All the ions may break apart and go their separate ways. One may coagulate the other at which point they will simply settle together to the surface and await further developments. One compound may chelate or sequester another who otherwise would not dissolve or move along. If the rain itself or the runoff thus impelled is particularly acidic, the running water may dissolve metals, bits of asphalt, or soil-ions that in more basic solutions would remain insoluble. Grass and street surfaces may for hours bleed the accumulated soluble hydrocarbons with which they have been bombarded for days.

Once in a sewer, still further reactions may occur or ones already started may persist toward completion or a new equilibrium. If the runoff reaches first an empty storm drain, it may remain largely unchanged in quality until it reaches the downstream lake or creek. If it falls among the sanitary wastes in a combined sewer, it surely will change, and be changed by, the largely organic, likely chemically different materials with which it mixes. The same kinds of changes can occur if it seeps through earth first and then infiltrates to what was designed to be, but only partially truly remains, a separate sanitary sewer.

As rainfall reaches its peak and runoff swells to its highest rates of flow, initial scouring and flushing of land and street surfaces will be over. Still the heaviest particles the storm can move will be moved now. Further dissolution of soil and other solid particles will continue. Sewers will be scoured of old, maybe encrusted, maybe putrid deposits. Reactions will continue apace. Infiltration to the earlier drier or even parched soil will abate as pores become glutted, and almost all the rain that falls will become runoff. Turbulence levels will be maximal and erosion of particles will be common wherever available soils and other particulates lie exposed.

At receiving waters, runoff sluiced through treatment devices of a conventional sort will have lost their heavier particles in man-made sedimentation basins, but large quantities of dissolved salts, minerals, metals, and nutrients will be discharged. Bottom-living and swimming or suspended forms of life will be subjected abruptly to a new chemical environment, and flow rates and velocities will increase with which they must cope. Attached forms will be buffeted. Some will be torn loose and swept away. Hardier individuals will endure.

Runoff not treated, discharged as an overflow of combined sewage bypassing treatment, or discharged as untreated storm drainage, will affect the stream similarly. But such untreated runoff will have a much larger load of silt or suspended solids of all description. It may be warmer than the stream, having been warmed by the surfaces over which it flowed, which in turn were warmed recently by the sun. Its dissolved organic load will be wholly unabated, and at some point downstream dissolved oxygen will be consumed by bacteria which will begin the degradation of the organics immediately upon their entry. Whether the depletion is critical to stream life depends on a number of factors. Silt in the flow will inexorably settle when the stream reaches a pool large enough, or until the flow subsides to a velocity slow enough, to permit this. Whether the settled silt and other solids alter the stream or its aquatic life forms adversely is problematical. Metals, organics, other cations and anions that have attached to the solids by adsorption or ion exchange will be released to the dissolved state again to rise in the stream's water column and have their own effects on wildlife, if the chemistry is right. That is, if acidic or anaerobic or other reducing conditions, or ion exchange should occur in the sediments, the once soluble materials now associated with the silt can easily become disassociated again and just as soluble as before; and given the right circumstances, they will.

As rain subsides on the urban surface, flows of runoff will begin to subside as well. Smaller and smaller particles will fall into the category of those too large to be transported. Dissolution of chemicals will occur at fewer and fewer surfaces. Finally the rain will stop altogether; runoff will stop shortly thereafter. Only pools and saturated ground will be left. The wind will continue to evaporate water, which it never quit doing, although its process was overwhelmed for a while. The soil will accept the remaining percolating water slowly but just as quickly as it can. The city will begin to dry. Chemicals remaining in the final drops will attain critical concentrations and begin to choose mating ionic partners with whom they will precipitate as salts and other compounds. Crystals will sparkle with still-damp newness. Salts will encrust on curbs from moist gray toward white. Silt and sand grains will dry and stir now only for the wind. The shake roof will dry and crack some more, making new sites to trap tomorrow's dew and wind-borne dust.

That's what happens when it rains on a city. But nobody ever proved it, scientifically and conclusively. The following subsections discuss several possible ways of going about that, each of which has the same level of monitoring requirements, but with its own philosophical bent or set of subobjectives.

Scientific Research

How could pure scientific research on urban runoff quality be accomplished? What is the minimum number of variables one would have to consider, that he would have to trap in sample bottles?

In the first place, a single urban area will not suffice. There are too many topographic, cultural, land development, and climatic variables involved to expect all possibilities to be covered in just one. Knoxville, Tennessee; New York City; South Lake Tahoe, California-Nevada; and Denver are quite different. They all have homes and businesses, paved areas and open lawns and parks, to be sure. But weather patterns, topography, traffic patterns, and cultural practices are widely different; so runoff quality, even in similar isolated storm events, could be expected to be quite different. Additionally, their types of receiving waters are quite different, so impacts could be expected to vary tremendously.

Within each urban area there are normally a number of subcatchments or watersheds, each having unique topography, land cover, and kinds and degrees of land use. Even soil types can be varied in localized areas. So areal distribution of sampling within a single urban area will be necessary.

Rain storms throughout a year can be quite different, not only in total quantity of rain, but in distribution of intensities within the storm period. Distribution and amount of rainfall can affect runoff quality. A gentle but protracted rain may never scour a large percentage of available, loose solids, simply because the particles are too heavy to be moved by the relatively slow rates of runoff that result. A short but intense rainfall may result in far greater pollutant loads to local receiving waters. Also, the antecedent moisture content of the soils and accumulated pollutants from dry-period build-up can vary tremendously and can directly affect runoff quality and amount. So a number of storms should be monitored in a given year.

Moreover, winter storms will create much different runoff characteristics if they fall on snow-covered or frozen but exposed surfaces than those to be experienced in runoff on unfrozen exposed lands. Likewise, summer thunderstorms will produce different results if they are widely scattered during a parching drought than if they are frequent. So it is probably necessary to sample over a period of years to discern these weather pattern effects.

Also, several times a year between storms it will be necessary to sample the dry land and paved surfaces to determine how, and how rapidly, potential pollutants are applied or accumulated on the surface. Effects of wind, natural degradation, and cultural practices such as street sweeping on dry-period distribution of potential pollutants should also be determined.

During storms and during dry periods a number of samples will be required at each site. Numerous runoff samples will be required during each storm simply to determine the total pollutant load washed-off, as well as its distribution throughout the storm. Dry periods will require multiple samples at or near a single location merely to permit analysis of dry solids of several sorts and some elutriated, washed-off, soluble materials that are accumulated on grasses, curbs, and other surfaces.

For each runoff sample (bottle of water), a number of laboratory analyses will have to be made for the concentration or amount of various physical properties or chemical constituents. Among these would be the total organic load (Chemical Oxygen Demand, COD, or the equivalent) and the dissolved and suspended solids content. Additionally several metals likely to be present, such as lead and zinc, should be measured. Several forms of nitrogen and phosphorus, which will act as algal and other plant nutrients in receiving waters, should be analyzed. An indicator microbial species, such as coliform organisms, should be monitored. And pH and dissolved oxygen should be measured routinely, since they are critical to reaction types and rates that may occur. Additionally, it is advisable that several complex organic compounds be measured. The pesticides are obvious examples.

Inhoffer (1979) has pointed out that there are already over 70,000 potential organic contaminants of both natural and man-made origin. In discussing that article, Sonnen (1979) has suggested that 2300 are probably worthy of consideration in runoff. There are now 133 "priority pollutants" which EPA believes are potentially hazardous and should be monitored. How many are actually measured routinely will largely become a matter of economics. For this work, it is suggested that 20 separate constituents would be an ambitious, if by no means complete, list of properties and would be revealing of the fundamental processes occurring.

Nonetheless, the scientific researcher in the end must define his particular subobjective; and there are numerous mechanisms and constituents from which to choose, and there are pieces of the entire urban runoff quality puzzle that are worthy of close scientific scrutiny by themselves. Among these mechanisms are the effects of rainfall itself, its own quality constituents, its impact energy as it affects dislodgement of particles on the ground, and its areal distribution. There are cultural patterns, such as the use of catch basins, the presence of cross-connections between sanitary and storm sewers, traffic patterns and speeds as they affect resuspension and blowing of particles away from streets and gutters, dry-period fallout and accumulation of air-borne pollutants on land surfaces, and the general level of littering and debris collection *vis a vis* street sweeping and generic degrees of pick up and cleanliness in watershed areas prompted by civic pride or the lack of it. Particular constituents or their interactions are potential research topics as well. The demand on dissolved oxygen exerted by certain inorganic constituents, such as iron, sulfur, and manganese, might be studied. Actual pathogens might be studied, rather than surrogates or indicator organisms such as coliforms. Partitioning of any number of constituents into dissolved and particulate fractions, volatile and nonvolatile solids, affinity for solids of different types and different particle sizes, the fate of petroleum hydrocarbons, and numerous other subtopics of constituent chemistry and physics are worth scientific attention.

Finally, it is for each researcher to determine that his lab work is to be performed under satisfactory levels of quality control for his monitoring and analysis equipment and quality assurance for the ways in which his samples are handled, prepared, and analyzed.

Building a Model

Since the 1960's a number of mathematical models of storm runoff and storm and combined sewage transport have been developed. The hydraulic portions of these models have been very useful and remarkably accurate, largely because they have been based on scientific principles of continuity and on well-verified hydrologic empiricisms. Many contain water quality portions as well, which are far less scientific and largely unproved. Their purpose has been to permit planning and preliminary design of facilities for controlling storm runoff and combined sewer overflows. Their use today is so commonplace that some investigators have suggested that the major purpose of water quality measurements nowadays is to calibrate these models. The argument they make is that the models contain all the latest theory of behavior; and expensive monitoring may be safely foregone, other than for a few sorties to grab some calibration data, because the models and computers can do all the deducing necessary.

Unfortunately, these investigators are wrong where urban runoff quality is concerned. Today's water quality models for storm runoff are numerically and algebraically intricate. Special jargon and on-the-job training are necessary to converse with them. As a result, a certain mystique surrounds their structure and functionings, and unfortunately this mystique if often translated to magical, powerful comprehensiveness. Models are touted as "the state-of-the-art" knowledge. College professors and renowned consultants are heralded for the discovery of these tools as though they represented quantum leaps forward in the insights and objectifications of man. This is a highly inaccurate picture; and the college professors and consultants involved know it.

The mathematical models, computer programs, available today for simulating runoff quality are astonishingly quick at performing voluminous arithmetic and printing of neat and perfect tables of values. They are precise to the point of lily-gilding. And they provide engineers easy and rapid opportunities for analyzing the hydraulic behavior and crude pollutant mass loadings that would result if one or another, or one of thousands of still other, alternative large-scale city-wide systems of engineered works such as sewers and plants were built. Hence, they are useful for preliminary engineering planning of alternative structures and land use patterns.

But, today's models are not scientifically thorough descriptions of runoff quality behavior. Indeed, they are extremely crude. They were made this way purposely. The purposes were: 1) to avoid field measurement of numerous verifying data, and 2) to achieve at least a body of sophisticated tools for preliminary planning even though long-term, reliable sets of field data had never been collected.

These same models, which will be discussed more fully under a preliminary design objective to which we have not yet gotten, are useful in that context. But hopefully enough has been said here to explode the young myth that they are statements of scientific principles of runoff constituent behavior. They are not.

However, a model of runoff behavior having scientific rigor and completeness would be an extremely useful tool. Most of the knowledge about physical-chemical processes necessary to construct such a model now exists. The following processes could be included: wetting and drying of surfaces, dissolution of soluble salts, crystallization of dissolved solids reaching the limits of their solubility, ion exchange/adsorption reactions, sedimentation and scour of particles, chelation and sequestering, oxidation-reduction reactions, and pH-dependency and temperature effects on reaction rates. Not only are these processes not represented holistically in any model existing today, but also not one, no single one, of these processes is represented in any urban runoff quality model today.

What would be required to derive and prove such a diversely useful tool in terms of calibration and verification data? Probably five years of monitoring in watersheds in at least three urban areas of the nation (or world). In short, precisely the same data gathering effort outlined before for a reasonably complete understanding of the processes involved. The beauty of the tool would be that enormous, continuing, comprehensive, monitoring programs in numerous other cities could then be foregone. Only a relatively few data would be necessary in other places for first-use calibration. But the tool would exist then, as it does not today, to predict behavior quickly and indoors by computer rather than through the laborious and time-consuming, outdoor expensive data gathering process. Unfortunately, data gathering cannot be foregone with impunity until this scientifically compelling and important step is taken.

OBJECTIVE II: PROBLEM IDENTIFICATION MONITORING

Land Runoff

It is probably clear by now that the major focal point of this work is on the quality of runoff from urban land surfaces, land runoff. For a host of reasons, investigators have studied and reported the results of monitoring in sewers, minor streams, waste treatment plant effluents, and major receiving waters. But the quality of storm runoff from lands and paved surfaces of various kinds, still on watersheds and prior to being mixed and integrated in sewers or streams with waters from many other sources, has not been thoroughly analyzed. And it is the runoff from these separate land sources that gives rise to the quality "problems" that runoff presents, if there are any. So we concentrate here on monitoring of quality in land runoff, although other subsections that follow discuss the special "problems" of discerning differences between point and nonpoint source loads and of determining the impacts of storm runoff on the beneficial uses of receiving waters.

But in Denver, for example, if we were compelled to determine if land runoff were causing a quality problem in our local streams, we would be faced with determining the source of such problems. In contrast to the monitoring just discussed for scientific research purposes, several important differences are obvious. First of all, our monitoring program would be confined to only one urban area, Denver, and not three or more as before. Other than the guidance provided by monitoring experience reported in the literature by New York or Miami or Des Moines, those cities' results would not be germane here, and certainly there would be no reason for our monitoring there. So the number of urban areas to sample is immediately reduced to one.

Secondly, since the question to be addressed is a fairly simple one: Do we have a problem or not; compared with the scientific, cause-and-effect, mechanistic questions posed in research; it seems clear that we could answer our question in a much shorter time and by sampling fewer storms than the researcher would require. We have only two types of storms here, summertime thunderstorms of highly intensive and highly localized downpours and wintertime frontal storms of lower intensity and more general geographic scope. In any reasonably "normal" year a number of both types will occur, so we are reasonably confident that we could determine whether urban runoff were a problem or not by sampling for a single year. Furthermore, since "sampling" *per se* is the objective here, i.e., getting only an indication of behavior from among the entire set of possible storm occurrences, as opposed to the scientific objective of more nearly characterizing the whole array of possibilities, it seems clear that we could sample far fewer storms. Five or six storms at each site appear quite adequate to determine if storm runoff at that place is potentially polluting or not.

Still, it is also true that, as for the scientific objective, we would need to monitor several watersheds within our area to distinguish among predominant land use types and land slopes, and we would likely need to sample at more than one site in each watershed. Moreover, as with the scientific objective, we would want to perform at least a modicum of dry-period sampling to distinguish occurrence and differences in pollutant build-up amounts and rates. So we would probably want to monitor in at least three watersheds, with at least two sites in each one; and we would sample at least twice in the area of each site during dry, interevent periods.

Quality constituents would be limited to those most often cited as potentially affecting beneficial uses. These would include a general organics indicator, such as BOD or COD, general solids indicators such as suspended or total solids and perhaps both, and algal nutrients such as nitrogen and phosphorus. Coliform organisms, and fecal coliforms in particular, would also probably be monitored to learn if potential bacterial contamination existed.

It is often argued that monitoring of land runoff in urban areas is ineffective and inconclusive because land uses change so rapidly and drastically, particularly with respect to the percentage of impervious area which has a major effect on runoff quantity and quality. This argument is largely spurious and specious. What it belies is merely that runoff monitoring has not normally been performed on small enough areas of sufficiently uniform land use types. Certainly monitoring results will change with time if samples are taken in a creek at the lower end of an agricultural or forested watershed now experiencing its first of many planned residential developments. And a partially completed neighborhood with many vacant lots, some disturbed by construction, will yield different runoff qualities than the completed development. But what is needed is a body of runoff quality data on open lands, forested lands, single-family homesites, parking lots, streets, and high-rise buildings and commercial and industrial areas. Then predictions could be made by aggregating the effects of further development of such individual units, and removal of others, over time. This would be the long-range intent of land runoff monitoring in our area. But a larger and more comprehensive monitoring program would be necessary than the one just outlined to answer merely if there were or were not a problem with runoff quality right now. Still, such monitoring could be much less ambitious than a comprehensive scientific inquiry.

Point vs Nonpoint Sources

In recent years, spurred by certain pollution control planning imperatives and existing ways and means for getting controls in place, the identification of whether or not a city had a problem with runoff quality has often been reduced to determining whether point or nonpoint sources of waters discharged to local receiving waters were the greater source of the potential pollutants. Point sources, such as waste treatment plant effluents, after all, are rather closely regulated; and it is reasonable to assume that if storm runoff is less noxious than these discharges, which meet certain agreed upon quality standards that ostensibly mark acceptably safe levels, then runoff must be safe as well, and even safer. Ergo, runoff is not a problem. Should the runoff constituents be greater in concentration and particularly in total mass than the discharges at point sources, then the argument of absolute safety is destroyed, but a number of options remain. The nonpoint sources could be controlled to the same or some other levels. The question could be defined as moot until some damage is tied to the runoff. Or, as a matter of policy, storm runoff could be defined as not constituting a problem since its causes are "natural," thereby obviating the need for any monitoring, which by this definition of "no-problem" could not prove otherwise. Each of these options has been invoked at various times and places.

But to determine the relative size of point vs nonpoint discharges, a monitoring program could be devised that would be slightly different and probably smaller than the land runoff program just described. Immediately it becomes obvious, for example, that no dry-period sampling would be necessary. Also the number of sites might be reduced, since the distinctions between and among land use types become less important. Still it might be necessary to monitor a larger number of watersheds, perhaps all of them, in a single urban area, but sampling at a single, downstream site near the receiving water should suffice. A number of storms would have to be sampled; and of course the point discharges would have to be monitored to permit the fundamental comparison, but presumably these are being sampled already for purposes of monitoring compliance with discharge requirements. It is to be noted, however, that monitoring of waste treatment plants is not always as frequent with respect to some constituents as this comparison may require. Also some plants are mysteriously badly sampled during storm events when they notoriously produce their worse-than-average effluent quality. So some additional monitoring of the point sources should be anticipated as well.

Comparisons should be possible of: the total annual load of various (say 5) constituents from each of the two source types; the concentration means, ranges and standard deviations from both sources; the frequency of violations of discharge standards for the point sources; the frequency of runoff events; the total volume of point and nonpoint discharges; and the relationship, or lack of one, between discharge violations at the point sources and the occurrence of runoff events.

Receiving Water Monitoring

It has been a fundamental theorem of pollution control law in America for decades (although it is sometimes overlooked) that pollution is defined by, it only exists when there is, the proved occurrence of damage to an instream or downstream beneficial user of a receiving water. Consequently, if Denver is to answer the question: Is runoff quality a problem; ultimately it will have to go to a receiving water and sample there to get the answer.

Receiving water monitoring has been performed for a variety of reasons for decades, and there are standard methods for performing these programs that are generically available. So we will not dwell on them here.

However, compared to sampling of runoff from land surfaces, even for determining the impacts of land runoff, receiving water sampling is much different and in some ways more difficult and expensive. There are equipment and paraphernalia differences, such as boats, bottom samplers, and SCUBA gear, that are obvious. Also there may be requirements for nets for algae and fish or microscope slides in diatometers that are both specialized in design and requiring specialized or even expert training for proper use. Still further, a single site in a river may require not one set of sample bottles per event but many sets so the differences in water quality and biotic effects can be discerned across the river at that single "point." (We find this particularly important in sampling streams in the Denver region.)

Perhaps most fundamentally, receiving water impact sampling differs from land runoff sampling in the things to be measured. Impacts in a receiving water are largely biological and to some degree stream-morphological, whereas runoff sampling is largely chemical. To be sure, the quality constituents dissolved in streams must be measured as well, and effects on the utility of the water and its safeness for downstream withdrawers must be inferred. But the most immediate and measurable, obvious impacts of runoff on receiving waters will be the impacts to the stream itself and the biological community(ies) that exist there. Consequently, receiving water sampling must consist largely of collection of samples of benthic organisms, fish, zooplankton and algae, and suspended and other present life forms, which can then be analyzed in the laboratory for the presence or absence of physical damage or monitored in the stream for short-term or prolonged stress. The causative elements, such as metal toxicity or smothering by solids, must also be identified; so many chemical properties of the water must also be measured. Although normally a measure of effects as much as causes, dissolved oxygen concentration is an obvious and important indicator of stressful conditions.

A 1979 conference sponsored by EPA [Yousef, *et al.*, 1980] on the effects of urban runoff on receiving water quality and its overall impacts on biota presented some results from a number of investigators. Largely these were university researchers whose objectives were primarily scientific in terms of rigor and specificity. That is, they went beyond asking whether or not a problem existed and addressed particular organisms, particular lakes or streams or estuaries, and particular cause-and-effect mechanisms. Despite that, on the whole the scientists reported uncovering more questions than answers; and generally their demonstrations of effects, which they often found, were nonetheless usually small or inconclusive. What they *seem* to have found, to date at least, is that *established* biological communities in receiving waters that have received storm runoff for eons are remarkably adaptable or adapted to transient inflows, even insults, from storm runoff. The *change* in biological quality owing to new development of once-open lands is a body of knowledge only now being explored. The requirements for long-term sampling to satisfy this objective can be appreciated.

Moreover, one can foresee that should the changes in biological communities in receiving waters be subtle ones, a new public debate will open on the question of whether those changes are indeed adverse or damaging. Biologists, who tend to value *balance* of communities more highly than absolute presence or absence of species (excepting gross pollution indicators such as mats of slimes or banks of sludge worms), will be of little help in resolving this political question.

Probably the most severe damages attributable to urban runoff pollution, which if found would justify its abatement measures and attendant costs, will prove to be health effects on man and interferences with or damages to industrial and commercial utility of the water withdrawn downstream. Unfortunately these damages, especially health and genetic damages, which may take generations to become evident, but which may be significantly large, will be difficult to quantify for decades to come.

OBJECTIVE III: ALTERNATIVE SOLUTION MONITORING

Again, under three subheadings, an equivalent monitoring program is described in the separate contexts of three preliminary design or planning subobjectives for selecting from alternative runoff quality control measures.

Best Management Practices

The Federally-mandated "208 planning process" mentioned earlier calls for determination of nonpoint source pollution problems and study to determine the "best management practice" that could be implemented to alleviate the problem elucidated through monitoring.

Largely because the Federal government, notably EPA, is stopped by policy from funding stormwater treatment works (i.e., separate storm water facilities), and because local governments are not compelled to provide these expensive measures, "best management practices" or BMP's have come to include less expensive, less intensive land treatments or runoff-intercepting devices. These include street sweeping, detention or "retention" basins to capture storm water for a while to permit suspended solids to settle, catch basin cleaning programs, and the like.

Monitoring to determine which of these alternative BMP's is indeed the best, in terms of technical effectiveness and cost, is slightly different from problem-identification monitoring and yet very much less demanding than scientific research. In older parlance, from days prior to "208 planning" and "BMP's," this level of monitoring is tantamount to "preliminary design" or "analysis of alternatives."

Again, only a single urban area is involved. Also samples of runoff from a variety of watersheds could be collected at a single downstream site in each watershed. A number of storms, roughly equivalent to those in the problem identification objective (5 or so) should be sampled throughout one year at each site. The constituents to be monitored would likely be very few, probably only solids or maybe solids and organic material (BOD or COD). The point of this monitoring is to determine the relative loads of material likely to be washed from each watershed (type) so appropriate BMP's could be applied.

It might be desirable, and some investigators have chosen to do this, to monitor the performance of a certain candidate BMP itself, already in operation. A good example is street sweeping. Sometimes, in cities where street sweeping is currently practiced for other than pollution control (amenity) reasons, sampling is done of runoff before and following a sweeping. In other cases, runoff is sampled from a basin or watershed or neighborhood receiving periodic sweeping, and the results are compared to runoff samples collected from a nearby control basin wherein sweeping is purposely foregone for a control period.

The advantage of directly monitoring a given practice is the clarity of the evidence that it works or that it does not, or just how well or poorly it works. The disadvantage, obviously, is that the BMP must already be in place, at least on a pilot scale.

Control of Combined Sewer Overflows

In many of the country's large and older cities, combined sewers continue to be in use and it is clear that they will remain in use for years to come. The storm-period overflows of combined sewage are a problem. Impacts from these flows have been demonstrated. Abatement of these problems has been slow in coming, and for a variety of reasons. As was indicated earlier, treatment plants to treat effectively the entire storm and sanitary flows in these sewers are prohibitively expensive. Moreover, states and the Federal government have been unable to establish meaningful discharge quality standards or receiving water standards for wet-weather periods, so expected performance levels, hence design criteria, have not been developed. Still, the quality of combined sewer overflows represents a clear and gross pollutional impact, and reasonable people agree that some further controls are necessary.

The dilemma posed by combined sewer overflows (CSO's) with poor quality and high treatment costs has largely led to the development of the concept of BMP's, cheaper and less intensive abatement practices to be instituted upstream on the various contributing watersheds. The concept very simply holds that downstream flows must be reduced through upstream detention storage, or quality of upstream flows must be improved (e.g., by street sweeping) so that the overflows downstream will be less polluting. Other alternatives for CSO control include: stormwater solids separators such as swirl or teacup concentrators, sewer flushing during dry periods to remove settled and accumulated silt and organic deposits, catch basin cleaning programs, and provision of partial downstream treatment, such as sedimentation or air flotation.

Monitoring to assist selection of one or more of these alternatives for implementation would be roughly equivalent to that just described for BMP selection. That is, in a single urban area, one or more sites in a number of watersheds (perhaps 10) would be sampled for roughly 5 storms per year. A small set of constituents could be analyzed. Some dry-period sampling for accumulation rates, on streets and parking lots or in sewers, would be desirable.

Mathematical Model Calibration

In the preliminary design, or evaluation of alternative solutions step toward understanding and controlling urban runoff pollution, mathematical models are being used more extensively than in any other step.

That is, this is the objective which they serve best and most often. Therefore, a few words about calibration of these models through monitoring is appropriate at this point.

As we have said before, there are numerous computer programs available today for performing hydrologic, hydraulic, and water quality computations. The programs or "models" described here are just that, models, or simulators that solve mathematical expressions for water flow or quality in somewhat to purely idealized, simplified representations of an urban watershed. These models consist of the algebraic equations and instructions for their stepwise solution, which comprises the "program," and they require a second component to function. This component is the input data.

Input data describe everything about the particular set of circumstances, the watershed or stream or sewer, being simulated; and they describe some of the processes taking place as well. For example, early in the program there may be a statement about the area of the watershed, such as

$$A = L \times B$$

where A is the area in square meters, L is the length of the (assumed rectangular) watershed in meters, and B is its width in meters. That expression could be used for any watershed. But to simulate a particular watershed, values of L and B, say 500 and 100 meters, must be given. This is one kind of data that we could go to the field and measure, or scale from a map; and we could feel confident about the accuracy and precision. Only the assumption about the rectangular shape would be open to question, and we could evaluate the precision of that assumption in each case. This kind of data gathering and use is not really calibration.

But there are other kinds of data these models require which make them function over a wide span of utility from very helpful to just plain wrong and misleading. For example, the better known models estimate concentrations of virtually all constituents in storm water as a direct function of the amount (concentration) of suspended solids. To wit,

$$PO_4^{-3} = EX \times SS$$

or

$$Cr^{+6} = WY \times SS$$

where SS is the suspended solids concentration, PO_4^{-3} is phosphate concentration, Cr^{+6} is chromium ion concentration, and EX and WY are coefficients (usually quite small) that are essentially decimal fractions or percentages of the suspended solids amounts present that are expected to be in the other chemical forms. So common has this practice become that Zison (1980) has written a whole book about determining values for these coefficients, which he calls "potency factors." (Zison's work also includes some interesting multiple correlation results with many sampled data from a variety of watersheds; and although the correlations obtained are for the most part not very good, which he acknowledges, nonetheless this work is well worth review.) But to calibrate a model of this type for a particular watershed one would have to sample in that watershed first, determine suspended solids masses and phosphate or chromium masses, and determine from the ratios of these what the appropriate values of EX and WY are. This is calibration.

Verification is the testing step beyond calibration, wherein the modeler uses the calibration values of EX and WY to simulate new storms; and he samples in the field again for each of the verification storms to see if the relationship between suspended solids and the other constituents holds up. Storm water modelers recalibrate and reverify their models of this type continually, everytime they simulate a new watershed. (At least, they should!) This is one evidence that such models are not very good representers of truth. Zison's (1980) poor correlation coefficients are another. Still such models are used broadly, and decisions are made on the basis of their results. They do permit simulation of many conditions, some of which do not exist yet and are hence incapable of being monitored in the field. They perform voluminous arithmetic speedily, and overall they are a convenience and generally worthwhile. They are not, however, correct. They are approximators, estimators, ways of making predictions based on little or no better information. They are not scientific advancements over monitored data; they are made and used explicitly and unabashedly to take the place of monitoring, specifically to avoid the scientific study of large cost and time-consuming labor outlined under the scientific objective.

There is another aspect of modeling, and particularly of modeling of urban runoff quality, that is worth exposing. Some models do contain statements of algebra purported by their developers to be science indeed, requiring no further calibration or verification. Some of these statements are completely unsupported by scientific data, and they are likely to lead to quite inappropriate conclusions. How damagingly wrong or inappropriate those conclusions might be would depend on specific circumstances. There is a glaring example that illustrates this point.

In several of the better known runoff quality models the relationship between the mass of pollutant remaining on a watershed after rainfall has started and the rate (flow rate) of runoff is

$$P_p(t + dt) = P_p(t)\exp -4.6Rdt$$

This is the models' representation of the washoff process for all constituents ostensibly, but really only for suspended solids, since the masses or concentrations of all other constituents are estimated from the solids masses via the potency factors, which are read as data. Now, this expression requires no calibration. It says that the mass of pollutant, P_p, at the next time step, t+dt, is equal to a function of the mass at the current time, t, the length of the time interval, dt, and the runoff rate, R, in inches per hour.

We shall not belabor the detailed points at which this expression is based on faulty or unverified assumptions, since Sonnen (1979) has explored these at length elsewhere. But he showed that: a) the expression is based on seven separate assumptions, probably none of which is true; b) washoff is actually a function of particle size, flow rate and depth and/or shear stress, the rate of rainfall, the degree of imperviousness and the roughness of the watershed surface, and its area and shape; and c) the value of 4.6, given as a constant, if more fully considered and developed from more exhaustive principles, could vary over three orders of magnitude (roughly 0.05 to 10) and moreover that 4.6 is very nearly at the high end of the range of plausible values.

In closing, we wish to reiterate that mathematical models are useful tools. They are used extensively today and they will be used even more widely in the future. They are not all subject to the same criticisms leveled here at a few or at parts of a few. Indeed the majority of runoff simulation models are excellent and highly useful at quantity, or hydrograph, simulation. The verification, repeatability, and accuracy of any number of hydrologic models is virtually awe-inspiring. They are phenomenally accurate at rainfall-to-runoff translation. Virtually all of them which include quality at all, however, are weak in this realm. The reason, of course, is that the theoretical statements to state quality behavior correctly or even adequately have simply not been deduced as yet, and that results directly from there never having been developed an adequate data base from which to derive those statements.

Nevertheless, today's models of quality of urban runoff can be calibrated; and the monitoring effort as outlined previously in this section is adequate to the task. The only expansion of such a program needed would be dependent on how many consitituents one would wish to simulate and hence on how many "potency factors" one would need to estimate from the measured data.

OBJECTIVE IV: DESIGN MONITORING

Perhaps the greatest distinction between monitoring objectives is that between design monitoring and scientific monitoring. This distinction is worthy of some consideration, since the required differences in sampling intensity and cost among objectives are so centrally to the

the thesis of this work.

Scientific monitoring, as was discussed earlier, addresses physical and chemical processes and leads to derivation of theorems and relationships about the joint behavior of many variables characterized by the measured data.

Design of physical works, such as stormwater detention ponds or treatment plants, is a process of sizing facilities to obtain arbitrary or otherwise pre-established performance levels. Hence design is normally performed by "plugging in" a few measured values from the field to "theoretical" statements of behavior for the chosen type of facility which were derived by earlier researchers and investigators from many field, laboratory, and pilot plant observations. As a very simple example, let us consider Mannings equation for flow in an open channel:

$$Q = \frac{1.486}{n} AR^{2/3}S^{1/2}$$

where Q = flow in cubic feet per second, A = cross-sectional area, R is the hydraulic radius in feet, and S is the longitudinal slope in feet per foot. The Manning n or roughness coefficient is essentially the only unknown in a conduit design problem, and it can be fairly accurately estimated by an experienced engineer through consultation of one of numerous available tabulations of this coefficient for a variety of pipe or channel surfaces. Chances are great that a design engineer would not monitor at all to verify his selection of an n value, although one or two sets of field measurements would suffice to do so. But Manning himself actually made thousands of measurements, and in sewers not open channels, to derive this expression originally; and other investigators working in laboratory flumes and other controlled conduits have duplicated his work.

Another example somewhat closer to the topic is the design of a sedimentation tank. The generally accepted "theory" of sedimentation tank design for waste treatment plants (American Society of Civil Engineers and Water Pollution Control Federation, 1959) has been a line-of-best-fit relationship between percentage removal of suspended solids and a computed surface overflow rate in units of volume/time/surface area of the tank. The data on which this relationship is based, and the relationship derived, are both given in the manual; and it takes even a casual observer little time to recognize that the numerous data used do not fit the derived relationship very well. Nonetheless, many tank designs have been made based on that relationship; and as with Mannings equation the designer has only one piece of data to measure in the field or to estimate, namely the influent suspended solids concentration. If he does not merely estimate that value from past experience, only a few grab samples from a local sewer will yield a usable average value. Lager, et al. (1977) have reported a similar relationship between suspended solids removal and surface loading rate for stormwaters, which might be consulted for design of detention basins; but the fit of the single relationship to the measured data is again not very good. Still the designer has merely to estimate or measure only one value, namely the influent suspended solids concentration. A manual by Hydroscience (1979) presents some slightly modified relationships between removals and rates, undoubtedly developed from a large data base, wherein different curves are given for different influent concentration levels (although the data themselves are not shown). Still, while the designer's chore has been presumably assisted by the ostensible improvement in the available "theory," his monitoring chore remains trivial -- he must obtain only a representative value for the influent suspended solids concentration.

So, the monitoring requirements to support design are generally quite small compared to the requirements for satisfying other objectives. First of all, only one urban area need be involved; only one watershed is relevant; and indeed only one site within the watershed is involved, namely the spot at which the sewer or pond or plant is to be located. Several storms, maybe as few as 2 or 3, need be monitored in a single year; and only the few design-controlling constituents need be monitored, such as suspended solids and BOD. In at least one storm and perhaps several, a stormwater facilities designer should monitor throughout a storm, taking perhaps ten samples, to determine variability of constituents throughout the storm so variations in behavior can be anticipated for low and high flow periods.

An important caution should be noted. These design monitoring guidelines are based on the assumption of traditional design practice for waste collection and treatment systems, wherein the designer relies on past theoretical developments of numerous scientific investigators. There has been a certain degree of that kind of background investigation, "theory" development, and pilot plant testing in the stormwater field as well. The works of Hydroscience (1979), Colston (1974), Lager, et al. (1977), Mein (1980), and Whipple (1979) indicate the degree of success thus far obtained in deriving theoretical design parameters and relationships. Overall the results remain rudimentary and preliminary. Some devices, such as swirl concentrators, have been characterized through research and pilot studies more thoroughly than others, such as detention basins. The hydraulic behavior of virtually all stormwater facilities remains far better understood than their quality behavior. Therefore, stormwater facilities designers should be cautioned that the monitoring requirements outlined here presume that design theory for the facility at hand is adequately developed. If that assumption is not valid, the designer may be faced with a monitoring and pilot program more nearly approaching that outlined earlier for the scientific objective which is far more ambitious. Whipple's article on detention basin design (Whipple, 1979) suggests that this will remain the case for some years in design of those particular facilities. Their quality behavior remains virtually unresolved.

OBJECTIVE V: REGULATORY COMPLIANCE MONITORING

To satisfy this objective, the sampling-program designer probably can see intuitively that he faces a monitoring program calling for a large number of samples.

First of all he is confronted with a program that must be continuous into the future. This kind of sampling is not to be performed once and then all the required information can be gleaned. Indeed the information sought is precisely the long-term presence or absence or exceedance of some property outside a specified range of acceptable levels. So the investigator, from the start, knows that his total number of samples over the period of sampling, perhaps 25 years or so, is going to be large.

However, specified regulatory limitations against which the sampler is comparing his periodic data lend a seeming sense of complete objectivity to this kind of sampling. It would seem that the objective is to sample from time to time for the specified test constituent, and if the measured level is below the mandated limit everything is fine, and if the measured level is greater than the limit a violation has been found. It seems tautologically simple.

From both regulatory and statistical standpoints, however, the seemingly obvious is far from factual. First of all, a discharger is likely to be confronted with several kinds of standards. There are those which state absolute limits not to be exceeded; and there are others which specify a maximum frequency of allowable exceedances, such as "coliform organisms shall not exceed 1,000/100 ml more than 5 percent of the time." As Ellis and Lacey (1980) have discussed in a very recent article, this latter type of standard is becoming more broadly used in a movement toward acceptance of the variability of both sampling accuracy and, more importantly, natural behavior. Still, as they point out further (Ellis and Lacey, 1980), considerable theoretical work in statistics needs to be done before meaningful acceptability criteria can be devised and meaningful, useful, truly protective but not unreasonable regulations can be written.

Haan (1977) and Lindgren and McElrath (1959) have described the statistical testing available thus far for acceptance sampling. The solutions quickly become complex for even the simplest cases, which are all the statisticians have mastered at this point, so there is little use to introduce here the spate of probability statements, hypothesis tests, binomial formulas, operating characteristics, Poisson approximations, and the like that have been advanced for testing the acceptability or rejectability of an endless conveyor belt of toothpaste tubes or bathroom scales. But the statistics books just referenced or virtually any other college text on the subject will surely contain enough on the subject for the interested reader. He should consult the index under "Type I and Type II errors."

Suffice it to say here that a Type I error is one in which a sample is taken and the measured result leads to rejection of the population (or "lot") when in fact the sample itself was an outlier and a violation did not, statistically speaking, occur. A Type II error is more or less the mirror image of that: A sample leads one to conclude that no violation is occurring when indeed in the remainder or another part of the population or discharge or "lot" there is or was or will be a violation. Type I errors, then, are rejections of a true hypothesis about acceptability, and Type II errors are acceptances of a false hypothesis. Each does occur with finite probabilities, and they are the very natural results of sampling discretely from a more-or-less continuous population.

Sampling for compliance can be contrasted with the scientific objective requirements to get an idea of the monitoring effort anticipated. In compliance monitoring there will obviously be only one urban area for the responsible party to treat, not three or more as in the scientific area. But the number of watersheds in the urban area could be much larger. This objective is not interested in merely sampling a few test watersheds; the law requires knowledge, proof, that there are no violations. Therefore all or a very large percentage of the watersheds discharging to regulated receiving waters will have to be monitored. As many as ten watersheds in a given urban area may have to be sampled, at least at one site per watershed near the point of stormwater discharge. This contrasts with the scientific objective in which several sites in each watershed could be monitored but under which only several, 2 or 3, watersheds need be monitored in each urban area.

In compliance monitoring most storms that occur will have to be sampled, certainly the larger and significant storm events. Twenty or so storm events per year are indicated. The number of samples per storm can be fewer than for the scientific objective, since each detail of the storm's time variability is not necessary -- the point is to sample that variability. But 3 to 5 samples per storm (as opposed to 15 or 20 in the scientific objective) will be required. Moreover, for some time to come, urban stormwater regulations will undoubtedly be limited to a fairly small number of constituents, probably always as many as 3 but rarely more than 5.

In compliance monitoring for runoff one notable advantage is the lack of any requirement for dry-period sampling.

OBJECTIVE VI: OPERATIONAL PERFORMANCE MONITORING

There are two subobjectives for monitoring urban runoff quality for long periods at a single site that are similar in design but different from the regulatory compliance monitoring just described. In monitoring performance of a treatment device or a detention basin, for example, the objective is merely to determine that the devices are working as they were designed to work and that they continue to work over the range of experienced variability of actual field conditions. So the subobjectives, which are described separately below but which have roughly equivalent requirements, are: Performance vs Design Criteria and Building a Record.

Performance vs Design Criteria

In monitoring performance of a treatment device or a BMP, there are a variety of measurements that could be made. The influent and effluent quality could be measured; interior qualities could be measured at various points in the device (such as the bottom material and the supernatant liquid in a detention basin); or only the effluent might be sampled. This will depend on the device being monitored, the design variables being checked, and the performance criteria originally specified (e.g., percent peak volume reduction, percent removal, or effluent concentration limits).

Generally speaking, though, these characteristics will be common. Sampling will be performed over a long period, probably the life of the device, say 25 years. Sampling will obviously take place in a single urban area, but at as many sites as there are devices to monitor. Probably only a few constituents will be analyzed, those that controlled the design. Sampling can be much more periodic than in monitoring for compliance with regulations. Probably three or four events per year will suffice, and three to five samples per storm should be sufficient. No dry period sampling is required unless the point is to determine sludge accumulation or a similar, particularly dry-period characteristic of performance.

Building a Record

Monitoring of operation of controls (or even of general behavior where controls are not in place) is similar to the Problem Identification monitoring described earlier. The difference is largely philosophical. In problem identification one is looking for problems; in operational monitoring one is looking for the absence of problems, but the requirements for sampling are roughly equivalent except that in operational monitoring the sampling continues for a long period.

But as conditions change over time, as new upstream lands develop, as more runoff is produced, or as unit loads of potential pollutants are increased, downstream controls may begin to fail or to fail more often than before. That is, where problems did not exist before, they may crop up. Having a continuous record that shows when the "problem" developed can materially aid in pinpointing where and why it arose so new controls can be imposed.

A record of long-term performance can be useful as well to document at least partially that newly found problems downstream of one's city were *not* caused by runoff from *this* source.

There are any number of reasons for building a long-term record of runoff control performance. Adding local knowledge to the engineering and planning professions is one. Protecting the city's interests in disputes with neighbors is another. Justification of expenditures for controls is a third. There are many more. As we shall see in the following chapter, operational performance monitoring is not cheap; it costs roughly $100,000 per year. But the information gained could be well worth the expenditure, in terms of saved costs for unncessary controls, in terms of improved confidence in designs for local controls that *are* put in place, and in terms of important evidence of nonculpability for presentation in disputed downstream claims or to support proposed regulations on upstream dischargers.

Again, the guidelines are: One urban area, a combination of approximately ten watersheds and sites, several storms per year and several samples per storm, with little or no dry-period sampling required.

REFERENCES FOR CHAPTER II

1. American Society of Civil Engineers and Water Pollution Control Federation, *Sewage Treatment Plant Design*, ASCE Manual of Practice No. 36, *WPCF Manual of Practice No. 8*, 1959, 357p.
2. Colston, N.V., Jr., *Characterization and Treatment of Urban Land Runoff*, EPA-670/2-74-096, National Environmental Research Center, Cincinnati, Ohio, December 1974, 157p.
3. Ellis, J.C., and R.F. Lacey, "Sampling: Defining the Task and Planning the Scheme," *Journal of the Institute of Water Pollution Control*, Vol. 79, No. 4, 1980, pp. 452-467.
4. Haan, Charles T., *Statistical Methods in Hydrology*, The Iowa State University Press, Ames, 1977, 378p.
5. Hydroscience, Inc., *A Statistical Method for Assessment of Urban Stormwater Loads - Impacts - Controls*, EPA 440/3-79-023, Prepared for EPA Non Point Sources Branch, Washington, D.C., May 1979, 338p.

6. Lager, J.A., W.G. Smith,, W.G. Lynard, R.M. Finn, and E.J. Finnemore, Urban Stormwater Management and Technology: Update and Users' Guide, EPA-600/8-77-014, Municipal Environmental Research Laboratory, Cincinnati, Ohio, September 1977, 312p.

7. Lindgren, B.W., and G.W. McElrath, Introduction to Probability and Statistics, The Macmillan Company, New York, 1959, 277p.

8. Mauzy, M., "Clean Water Act's Precepts to Endure," Water & Sewage Works, October 1980, pp. 38-40.

9. Mein, R.G., "Analysis of Detention Basin Systems," Water Resources Bulletin, AWRA, Vol. 16, No. 5, October 1980, pp. 824-829.

10. Sonnen, M.B., "Urban Runoff Quality: Information Needs," Journal of the Technical Councils, ASCE, Vol. 106, No. TC1, August 1980, pp. 29-40.

11. Whipple, W., Jr., "Dual-Purpose Detention Basins," Journal of the Water Resources Planning and Management Division, ASCE, Vol. 105, No. WR2, September 1979, pp. 403-412.

12. Yousef, Y.A., M.P. Wanielista, W.M. McLellon, and J.S. Taylor, editors, Urban Stormwater and Combined Sewer Overflow Impact on Receiving Water Bodies, EPA 600/9-80-056, Proceedings of the National Conference, Orlando, Florida, November 26-28, 1979, Municipal Environmental Research Laboratory, Cincinnati, Ohio, December 1980, 663p.

13. Zison, S.W., Sediment-Pollutant Relationships in Runoff from Selected Agricultural, Suburban, and Urban Watersheds: A Statistical Correlation Study, EPA-600/3-80-022, EPA Environmental Research Laboratory, Athens, Georgia, January 1980, 135p.

III. PREVIOUS MONITORING PROGRAMS

DENVER

Introduction

In February 1977 the Urban Drainage and Flood Control District (the District), the Denver Water Department and the United States Geological Survey (USGS) entered into an agreement for the USGS to perform an urban runoff quality study. A memorandum of understanding, as part of this agreement, contained the following five objectives:

1. Provide data to improve knowledge of the types, concentration, and behavior of dissolved and suspended material in storm water runoff in the Denver Metropolitan Area.
2. Provide information to evaluate the need for treatment, the technical effectiveness of various types and degrees of treatment, and the cost effectiveness of treatment. Possible degrees of treatment to be studied might include those to meet either stream standards and/or municipal water supply standards.
3. To the extent practicable the study should provide data to calibrate storm water quality and quantity models such as the EPA Storm Water Management Model.
4. Evaluate the impact of urban storm water runoff on the quality of: a) the South Platte River by comparison of storm runoff quality with the quality of South Platte River at Henderson, Colorado; b) principal tributaries of the South Platte--Clear Creek, Bear Creek, and Sand Creek; and c) the minor gulches in the area such as Sanderson and Weir Gulch, provided that baseline information describing the quality of water in the gulches were available.
5. Provide data on both the quantity and quality of storm runoff as a basis for more complete study of the feasibility of the use of urban storm runoff as a potential municipal water supply.

The program was funded by the District, the Denver Water Department (DWD), the Denver Regional Council of Governments (DRCOG) -- as part of their 208 effort, and USGS.

The final cost breakdown among the project sponsors was as follows:

District	$ 32,143
DWD	32,143
DRCOG	23,254
USGS	87,540
	$175,080

In an effort to meet the stated objectives, USGS established three water quality monitoring stations. These stations collected urban runoff from three separate urban basins in the Denver area. Data acquisition also included 5-minute rainfall depths at up to 3 gages per basin and runoff flow rates. The data were subsequently reported in USGS Open File Report 78-410, by S. Ellis, "Hydrologic Data For Urban Storm Runoff From Three Localities In the Denver Metropolitan Area, Colorado," May 1978. Following the Open File Report the USGS released an interpretive report by Sherman R. Ellis and William M. Alley, "Quantity and Quality of Urban Runoff from Three Localities for the Denver Metropolitan Area, Colorado," Water-Resources Investigations 79-64, May 1979. The reader may consult these two documents for specific information about the study basins, the data that were collected, and the interpretations and conclusions by USGS that resulted from the analysis of the data.

Special Considerations for Stormwater Data Collection

Although not mentioned in the interpretive report, collection of urban runoff data in a semi-arid climate, such as that in the Denver area, requires considerable care and effort. Most rainfall occurs as fast-moving thunderstorms, normally short-lived, and limited in the area they cover. For a majority of the runoff events it is virtually impossible for a sampling person to get to the site after rainfall starts and while runoff is still occurring. Automated flow measuring and sampling must be used; and in this case, they were used. Since the likelihood of runoff-producing storms is limited in a semi-arid climate, the automated equipment needs to be maintained in good working condition at all times. One has only a few chances to sample runoff; therefore a missed runoff event caused by equipment failure can be a major loss. As a result, on a per sample basis data will be much more expensive to collect. The automatic equipment will require more attention, and manpower requirements may have to be greater than in regions having more reliable and plentiful rainfall. Nevertheless, it is clear that the meteorologic character of a region needs to be considered when designing a runoff data collection program.

In a semi-arid climate special consideration must be given to the spatial or areal distribution of rainfall. When rainfall data are to be collected with the intent of relating rainfall to runoff, the raingage network needs to account for the spatial distribution of rainfall. In one of the basins of the Denver program, three five-minute recording rain gages per square mile were used. Afterwards, when USGS was calibrating the EPA's SWMM model, it became obvious that a single raingage would have been totally inadequate to model the recorded hydrographs.

Examination of Project Objectives and Program Achievements

Let us examine the 1977 program objectives, the data collection effort, and the reported conclusions. The first objective in the memo of understanding for the study was

> "To provide data to improve knowledge of the types, concentration, and behavior of dissolved and suspended material in storm water runoff in the Denver Metropolitan Area."

Data were collected at the three instrumented sites. They were analyzed and interpreted. It is safe to say this objective was generally achieved. Knowledge of the types and concentrations of dissolved and suspended material in storm water was improved for the Denver area. On the other hand, nothing was answered as to the behavior of the dissolved and suspended materials in storm water. Nevertheless, progressing from no data at all to a limited data base of reliable accuracy is a significant step forward.

The second objective was:

> "To provide information to evaluate the need for treatment, the effectiveness of various types and degrees of treatment, and the cost effectiveness of treatment. Possible degrees of treatment might include those to meet stream standards and/or municipal water supply standards."

Outside of some comparisons of the storm water quality against water quality standards and against the quality of secondary effluents, the USGS interpretive report contains little information that could be used "to evaluate the need, types, effectiveness, and cost effectiveness of storm water treatment." An attempt was made by USGS to obtain information on the frequency of street sweepings, but the data did not permit definite conclusions, and USGS clearly recognized and stated this finding regarding its effectiveness. The data did contain information on the dissolved and suspended constituents in storm water, but the data were not adequate to tell us if storm water treatment by settlement or filtering would be effective. For these reasons, it is felt that the second objective was not met, nor was it possible to be met with the data collected during the 1977 project. It may be that the total project budget did not anticipate the considerable complexities of this objective, and it may have been unrealistic to even attempt it within the available budget. Thus, this objective may have been sacrificed or dropped from consideration near the outset; however, no mention of this is made in the interpretive report. The interpretive report would have been more valuable if it discussed the original project objectives (which it does not), if and when they were modified during the project (which it does not), and what the reasons were, if any, for modifying any of the original project objectives (which it does not address).

The third objective stated:

> "To the extent practicable the study will provide data to calibrate storm water quality and quantity models such as the Storm Water Management Model."

It is virtually impossible not to meet an objective that begins with, "To the extent practicable..." Data were provided from three separate catchments. An attempt was made to calibrate the SWMM model for two of the sites. From the description of the calibrating procedure in the interpretive report, it appears that parameters such as percent imperviousness, and total basin area were increased significantly from the values measured in the field. Although reasonable flow simulation was achieved of the measured events, these events were very small (0.02 to 0.13 cfs/acre), and the excessive manipulation of basin area and percent impervious parameters makes the model's flow calibration suspect. This is especially true if rainfall events larger than measured ones were to be used in simulation during subsequent studies. Only one water quality constituent was modelled, namely, total nitrogen. On the basis of this modeling effort, Ellis and Alley indicated that if the initial buildup load were calibrated, the SWMM model (Version II) would reasonably simulate the total nitrogen washoff rates during a rainstorm. On the other hand, they concluded that the use of SWMM default linear buildup rates incorporated in the model overestimated total nitrogen loads by factors as large as ten.

Obviously an attempt was made to meet the third objective and to a degree, it was met. The two USGS reports do not indicate if the two test basin models were set up early in the program with the intent of meeting the stated objective. Looking at the "calibration" and "verification" comparisons in these reports, one is left with the impression that modeling came near the end of the data collection program. Lack of early commitment of resources to an objective or taking a "we can wait until data come in" attitude can result in a failure to meet any objective. If this happens, success is often marginal at best and is, to a large extent, dependent on luck. It is hard to say if the modeling objective would have been more successful had it been designed and pursued from the very beginning of the data collection effort. At least the investigators would have been in a position to recognize shortcomings in the data at the time they were being collected and then to report the pitfalls, problems, shortcomings, and successes encountered.

The fourth objective of the 1977 program was:

> "Evaluate the impact of urban storm water runoff on the quality of: a) the South Platte River by comparison of storm runoff quality with the quality of the South Platte River at Henderson, Colorado; b) the principal tributaries of the South Platte - Clear Creek, Bear Creek, and Sand Creek; and c) the minor gulches in the area such as Sanderson and Weir Gulch, provided baseline information describing the quality of water in the gulches is available."

Again, in light of the available budget, meeting this objective may not have been possible. If comparison of storm water runoff data and unrelated stream water quality data were the original intent of the fourth objective, then the program may have achieved it. On the other hand if the intent of the objective were to establish a link between urban runoff and stream quality, then it was not achieved. The interpretive report does have a section entitled, "Comparison With Water Quality of Local Streams" which draws a loosely worded conclusion that "...storm runoff may be a significant contributor of total ammonia nitrogen to local receiving waters. In addition, during winter months, snowmelt runoff may be a significant contributor of sodium and chloride to local receiving waters." (Emphasis added)

It is interesting to note that "significant" was never defined, not even as an order of magnitude. These conclusions were based on the data collected by USGS at the three study catchments and on a number of grab samples of stream water collected by the Colorado Department of Health in a totally unrelated program. The stream water quality samples were taken mostly during periods of good weather without any concurrent flow measurement. Comparisons were made on the basis of the water quality data obtained from the stream and from the storm water runoff program for a number of constituents including specific conductance, total ammonia, sodium, and chloride. Although data were also analyzed for pH, sulfate, nitrite plus nitrate, Kjeldahl nitrogen, total nonfiltrable residue, arsenic, copper, iron, lead and zinc; comparisons of river water quality with urban storm runoff quality for those constituents were not made in the USGS report for the following reason quoted in the report: "The other water-quality constituents were associated more with stream sediments than with the water column or were considered of lesser importance, based on the comparisons with the water quality standards and with secondary-wastewater treatment-plant effluent." The dropping from consideration of these "other" constituents for the stated reasons is in itself amazing.

The program approach, monitoring network design, and investigations performed would not permit this study to meet objective four. Had all the data acquisition needs and the analyses required to meet objective four been considered at the start, it probably would have been obvious to the investigators that an objective to evaluate the impact of urban storm water runoff on the quality of the South Platte River and its tributaries could not be met within the study budget. Readjustment or dropping of the objective would have been an appropriate decision and action to be taken early in the project. Instead, the study attempted to salvage it by drawing limited conclusions on the basis of a monitoring program result that in no way was designed to answer the key questions stated in objective four. The conclusions in the interpretive report do not appear to be substantiated by the data base or its analyses. Nevertheless, the conclusions leave the reader with an impression that the data do substantiate them. Unfortunately, such conclusions can mislead decision-makers, environmental groups and politicians into demanding costly and maybe even unnecessary control measures or into ignoring justifiable control measures.

The fifth and final objective of the Denver program was:

> "To provide data on both quantity and quality of storm runoff as a basis for more complete study of the feasibility of the use of urban storm runoff as a potential municipal water supply."

The interpretive report contains a discussion comparing some of the constituents found in urban runoff to drinking water quality standards. The basic intent of providing data for a "more complete study" appears to have been satisfied. This is particularly true when one considers there was a dearth of such data in the Denver area prior to the 1977 program.

Observations Regarding the Wet Weather Monitoring Program in Denver

The urban runoff quality study conducted by USGS in 1977 and 1978 was a major first step in collecting urban runoff quality data for the Denver metropolitan area. The effort partially filled a gap in such information for the semi-arid Rocky Mountain Region of the United States. The program, as originally conceived, was very ambitious with commendable goals and objectives. This program, like many others across the United States, underrated the complexities and fiscal and manpower needs to meet these objectives. This program, like others, fell into a pattern of collecting data without examining or reexamining the objectives. It is not clear from the reports if the objectives were fully considered in the design of the data collection network and in the selection of analytical needs and/or techniques required to meet the originally stated objectives.

It is also important to recognize the data collected during the 1977 Denver study are valuable. The data collection effort was one of the best. The water sample handling techniques, laboratory analyses and reporting can be used as examples for others to follow. However, as strong and good as the actual data gathering effort of the program was, collecting of data in itself is rarely the sole objective of any program and it was not the sole objective of this program.

It is easy to look back from today's level of knowledge and to point to the shortcomings of past efforts. On the other hand, we learn from the past. It is not the intent to criticize the 1977 Denver effort in the study of urban runoff quality. In fact it was a commendable and worthwhile undertaking. We are being penetrating in the evaluation of initial goals and final achievements to illustrate the pitfalls of setting objectives and not following through in the design of the data collection and data analysis efforts. Design of a storm water runoff study effort, including data collection, should be done with the objectives in mind. Even more important is the establishment of realistic objectives. This has to be done with full commitment of the resources needed to meet them once they are set.

HOUSTON

A fairly recent report (Characklis, et al., 1979) describes an extensive and ambitious urban runoff monitoring program which had very definite and lofty objectives but which appears to have failed to satisfy those, while it led its authors to draw a number of generalized conclusions about the behavior of urban areas which were not the originally stated objectives, and which the reported results do not support. This report is hardly a curiosity or a unique example; it is merely one of the latest.

According to Characklis, et al., (1979) The Woodlands "planned community" near Houston, Texas, was to have a thoroughly planned and controlled hydrologic system including provision for both flood control and drainage of local runoff and for storage and subsequent innocuous uses of the collected water. The stated objectives of their monitoring and analysis study, (Characklis, et al., 1979) therefore, were:

1. "...to modify and expand the capabilities of the EPA Storm Water Management Model (SWMM) to apply to the 'natural' drainage system designed for The Woodlands," and
2) "...determining effects of land use on storm water quality..."

The authors report:

> In order to accomplish these two primary objectives, a massive sampling and monitoring program was established. Rainfall, streamflow and over twenty-five water quality parameters were monitored on a regular basis. Included in the water quality program were intensive programs concentrating on bacteriological water quality, chlorinated hydrocarbons and phyto-

> plankton identification and enumeration. (Characklis, et al., 1979, emphasis added.)

These authors stated two objectives, and they went to the field and gathered data, which they report in profuse quantities, for at least 17 separate storm events over a 2-1/2 year period, with as many as four sampling sites being monitored during each storm.

Here is an annotation of their conclusions with respect to their first objective: modification of the SWMM model. Dr. Sonnen has marked with an asterisk those conclusions that appear to be supported by the reported data.

*1) After modification, SWMM could make "relatively accurate" predictions of peak and total discharge, but pollutant concentrations could not be modeled.

2) Modeling of erosion from construction activities is now possible with the modified SWMM version.

*3) BOD modeling proved unsatisfactory, because BOD monitoring was unsatisfactory. COD data were better and were used to model oxygen demand.

*4) Linear equations resulted to relate loading rates and total pollutant transport from a watershed.

5) "The exponential pollutant removal or decay coefficient can be considered as a constant in all geographical areas." SWMM was modified to accept a value for this coefficient selected by the user.

*6) The modified SWMM describes each natural channel with a series of coordinates and calculates stage-discharge curves.

7) Given specific local cost data or resorting to "default" values built into the model, the new SWMM can "determine the cost efficiencies in use of natural drainage systems" as opposed to more conventional pipe systems.

*8) The modified model can provide a detailed water budget for porous pavement systems, but it cannot do that for quality.

*9) Model modifications made in the study "require considerable input data preparation and, consequently, the modified SWMM, when applied to natural drainage systems, is more user dependent" [than the previous versions?]

The eleven conclusions reached about the second objective, determining the effects of land use on water quality, are similarly annotated as follows:

1) Stormwater runoff from undeveloped, forested land is relatively low in "pollutants and pollutant indicators."

2) Development in forested watersheds significantly increases suspended solids and nutrient concentrations in runoff. A number of related generalizations are also concluded.

*3) Urban runoff contains higher nutrient and solids concentrations than does forest-land runoff.

*4) One man-made lake served to trap "excessive" sediments transported by construction-site runoff. An 81 percent reduction was noted.

5) "A definite first flush is observed for urban and undeveloped watershed runoff..." "Urban drainage systems have increased transport potential and therefore exhibit higher flush concentrations." [These are not supported by the data given, and the Denver experiences in this regard have been quite different. Also, the potential for efficient transport means very little if there is no material available to be transported.]

*6) Rainwater contains phosphate, nitrogen, and COD...

6a) "...which account for a significant portion of runoff pollutant loads."

7) Undisturbed soils are capable of removing nutrients, while disturbed soils are not.

*8) Nitrogen and phosphorus in urban runoff tend to be in lower concentrations than in secondary municipal effluent; total suspended solids and COD concentrations are roughly equivalent to those in effluent.

9) "A linear relationship exists between total pollutant loads and total stormwater runoff..." [A tenuous generalization, at best.]

*10) On the basis of pounds per acre per inch of runoff, urban areas are substantially higher producers of suspended solids and nutrients in runoff than are forested watersheds.

*11) Simulation modeling [thoughtfully calibrated and applied]

> can produce "reasonable comparisons of observed and simulated curves."

Our conclusions about this study are, and again this is just one example of what appears to be happening over and over again in these urban runoff monitoring programs: The data collected are not well tied to the stated objectives of the study, and there is rarely enough time [money] to gather sufficient data bases to support statistically defensible generalizations, *especially* with regard to basic cause-effect relationships.

CHICAGO

As part of a 208 program in northeastern Illinois, two water quality sampling programs were undertaken in the past 5 years (1975-1980). The designated agency for performing the 208 studies in this region was the Northeastern Illinois Planning Commission (NIPC). The Metropolitan Sanitary District of Greater Chicago (MSDGC) performed the water quality sampling and analysis portion of the study.

Two separate programs were conducted. One was a study of the quality of runoff from watersheds having homogeneous, singular land use types, including various urban land uses. The second study was directed at characterization of quality in local receiving streams during and following storm events. Both studies have been summarized by Polls, *et al*. (1978).

Land Use Runoff Sampling Program

The *objective* for the sampling of runoff from urban and other land uses in Chicago and the surrounding region has been given by Polls (December 1978):

> The purpose of the land use runoff sampling program was to characterize the quality of stormwater runoff from sixteen small watersheds of homogeneous land use. By means of a water quality model, the data [were to be] used to predict the quality of runoff from similar sites in the six-county area to facilitate planning and decisions concerning treatment methodologies and management strategies for controlling non-point pollution.

It is to be noted that there were actually several, separable objectives, then. The first was to characterize runoff behavior from land uses of various types. The second was to calibrate a mathematical model, the specific name and type of which is not important, but which assuredly expressed behavior in certain fixed mathematical statements and simplifying assumptions. Then a third objective, ostensibly tied to the second but not the first, was to assist the planning for treatment and other strategies for controlling the runoff qualities found.

Polls (December 1978) has described the sampling program and presented the collected data. Polls and Lanyon (1980) have presented some interpretation of the results, although curiously nothing was reported about the modeling that was to be done or the degree to which the data collected were useful in calibration, and nothing is reported on the degree to which treatment strategies were planned or obviated on the basis of the data collected.

Sixteen watersheds with one sampling station in each one were selected. The watersheds totaled 36.24 square miles in area, which were used to characterize the 3,740 square miles in the six-county region. Agricultural watersheds comprised 22.67 square miles and residential lands made up 7.18 square miles. These areas predominated in the study because agricultural lands predominate in the region geographically and because residential lands are the fastest growing land use in the region (Polls, December 1978). Commercial, light industrial, highway, and forest lands were the other land uses sampled.

It is interesting to note that selection of the sample watersheds and sampling locations consumed 11 months, while the actual sampling of 86 storm events at the 16 sites took less than four months (April to July, 1977). The criteria for site selection, which appear to be logical and fairly simple to meet, were these:

1. Homogeneity of a single land use in each watershed.

2. Storm runoff only contributing from upstream (no combined sewage or separate sanitary sewage.)

3. Flow determination possible at or near the sampling site.

4. Electrical service available.

5. Suitability for sampler housing and reasonable freedom from vandalism.

6. Safety and physical accessibility for sampling personnel.

7. Land rights-of-way available.

To meet these criteria and to select 16 sites, the sample planners had to evaluate 100 sites, eliminating 84; and, again, the process took eleven months. (They had planned a three-month site-selection process.) Sites were eliminated by failures to meet any of the seven criteria, and all seven were violated at least once at the 84 stations dropped. Major stumbling blocks included finding that land use was not as homegeneous and singular as it had appeared from reference to maps, which was expectable, and failure to obtain rights of entry to the sites, particularly from other governmental agencies, which came as a bit of a surprise.

Samples were collected automatically in 1 gallon bottles. Field personnel were dispatched to the sites immediately upon receipt of telemetered rainfall data. An "acceptable event" was defined as one in which all sampling equipment operated satisfactorily, at least six water samples were collected during a storm, and the occurrence of rainfall and runoff constituted a discrete event. At least four "acceptable events" were planned to be sampled at each site, but less than four were sampled at three sites for a variety of reasons, including logistics, localized hydrology, and vandalism. A typical sampling event took approximately five to nine hours from the time the crew members were first notified of rain until they delivered the iced samples to the laboratory.

The quality constituents measured included oxygen demand: 5-day and 20-day BOD's plus COD; solids: total, suspended, and volatile suspended; nutrients; total kjeldahl, ammonium, nitrate-nitrite, soluble phosphorus, and pH; and metals: lead, zinc, copper, mercury, and chromium. The metal data have not as yet been reported or interpreted (Polls and Lanyon, 1980). Moreover, the COD data given by Polls and Lanyon (1980) are not reported in the more basic data book, giving results by station, as presented by Polls (December 1978).

Conclusions drawn by Polls and Lanyon (1980) from the measured data are fairly general; and they caution that, "Detailed comparisons cannot be made due to the large number of variables associated with each watershed and storm event." Nonetheless, they concluded:

1. Commercial land produced the greatest concentrations of pollutants, and forested lands produced the least, generally speaking with regard to all constituents as a whole.

2. Greatest variations in concentrations at single sites, caused by differences in storm events, antecedent conditions, and other factors, were among the total and suspended solids species.

3. Oxygen demand concentrations did not vary tremendously among land uses. [The mean 5-day BOD concentrations varied from 13 to 22 mg/l on six of the land use types. The value for forest land (for which only one storm was sampled) was 2 mg/l. The 20-day BOD range in mean concentrations was 23 to 55 mg/l. The COD values ranged from 78 to 168. The highest concentrations of organics were found to occur in single-family residential areas: BOD-5, 80 mg/l; BOD-20, 132 mg/l; COD, 308 mg/l.]

4. "...the mean values for nutrients reported for most storm events exceed the 0.3 mg/l inorganic nitrogen and 0.01 mg/l total phosphate suggested as threshold levels for algal blooms."

5. With relation to sanitary sewage in three waste treatment plants in the area, most constituents were found to be lower in runoff than in untreated sewage, but many, such as BOD and COD, were equal to or greater in concentration than they were found to be in secondary effluent.

The final, and telling, conclusion by Polls and Lanyon (1980) was:

> Further studies are needed to define the characteristics and sampling frequency necessary to obtain a statistically reliable data base for accurately determining nonpoint-source pollutant concentrations and loads from homogeneous land and watersheds.

Instream Sampling Program

As a part of the same 208 study, water quality data were collected in 1976 and 1977 at 45 stations on a number of rivers and streams in the six-county region. The purpose or objective of the program was to calibrate a mathematical model, and in this case planning and decision-making were only implied and not explicitly stated as corollary or subordinate objectives (Polls, July 1978).

The 45 stations were located on the mainstems or branches of 23 rivers and creeks. The criteria for selecting the sites were:

1. A desire to sample both the main river or stream and its tributaries.

2. A desire to sample both upstream and downstream of major point sources of pollution.

3. A need to monitor near junctions of major waterways.

4. A desire to sample near interstate boundaries.

5. A desire to measure in areas of major environmental concern.

6. A need to monitor at points critical to the behavior of specific constituents.

7. Safety and accessibility to the sites.

8. A need for nearby flow determination.

"Several" hundred sites were first located as candidates, and these were subsequently reduced to the 45, principally for time and cost constraining reasons.

Because response times were expected to be over several hours, it was determined that manual sampling would be preferable to sampling with automated equipment. Sampling was performed at 8-hour intervals for 10-day periods, although reductions in this frequency were made after an initial period. The reason, not given by Polls (July 1978) was apparently budgetary.

Two or three sampling crews of two individuals each in radio-equipped vans sampled for an 8-hour period daily, collecting samples three times a day at various stations in their region. The same stations were sampled for up to 10 days. Consequently, each of the three major drainage basins was sampled in one month, and all three were sampled in each 3-month season of the year. Sampling continued from April 1976 until March 1977.

In addition to the constituents reported for the land use monitoring program, analyses were also made for alkalinity, hardness, silica, chlorophyll a, total organic carbon, and turbidity. Lead, zinc, copper, chromium, and mercury were not analyzed in this program.

No conclusions have been drawn from these data by either Polls (July 1978) or by Polls and Lanyon (1980). It is not possible to report here how useful the data have been in calibrating a mathematical simulation model for river or stream quality. However, no rainfall data are reported in the data volume (Polls, July 1978), so correlations or correspondences of the measured data with runoff events would have to be deduced from data from other sources.

SAN FRANCISCO*

In 1975 the Association of Bay Area Governments (ABAG) was named the 208 planning agency for the San Francisco Bay Area, comprised of nine counties. Eventual funding for a total program to produce an Environmental Management Plan, consisting of water quality, air quality, solid waste, and other elements, amounted to 4.3 million dollars. The Surface Runoff Management Program was the largest single element with a budget of more than a million dollars. Of these funds, $160,000 were devoted to field sampling and analysis of surface runoff quality data.

---

*Material in this section was provided by Dr. Yoram Litwin of RAMLIT Associates, Berkeley, California.

The sampling program had several stated objectives:

1. To obtain information on the relationship between land use and storm runoff water quality. Specifically, data from the sampling program were to be used to estimate pollutant mass emission rates from various watersheds with different types of land use.

2. To use the collected data to calibrate and verify stormwater quality models that could then be used to estimate loads for the balance of the unsampled region.

3. To demonstrate conclusively the existence of certain runoff generated water quality problems which could not easily be dismissed by local decision makers.

4. To establish a basis for a continuing sampling network in the region which would develop data about the long-term trends of runoff quality and which would permit evaluation of water quality before and after implementation of the Environmental Management Plan.

An interesting feature of this program was its performance by the individual counties. Over 90 percent of ABAG's program funds were transferred to the nine counties for sample program planning. ABAG then provided a number of services to assist the counties and to maximize overall continuity of results. The Association was in charge of all the mathematical modeling; it coordinated and assisted in execution of the sampling programs; it provided overall guidance on individual technical issues such as projections of land use and population; and it developed a regional summary of the results from the individual programs. Additionally it provided the counties with written technical guidelines for sampling; it distributed a number of technical documents on the subject prepared by others, such as EPA and the USGS; and it conducted two technical workshops for training county staff personnel and included one field demonstration of sampling techniques.

In the end, 24 watersheds were monitored, and 55 storm events were sampled between November 1976 and April 1977. Useful data were produced; and ABAG was generally very pleased with the program.

Still, there were major shortcomings. These largely resulted from ABAG's ambitious desire to have the nine counties each participate directly. Viewed from hindsight, the difficulties that ensued seem predictable. First, some counties got a late start because the paperwork involved with approving and signing contracts with ABAG was delayed; and many did not get underway until January, having then missed a substantial portion of the area's rainy season. Certain tasks and general transmission of information to ABAG were performed slowly and beyond established deadlines, largely because county staffs were busy with other duties. It was also difficult to attain consensus amoung the counties on sampling techniques, constituents to be monitored, criteria for sample-site selection, and choice of analytical labs to perform the laboratory work. The obvious result was: These things were not done uniformly. Also some counties chose ultimately to subcontract the whole program to consultants, which obviated much of ABAG's efforts to involve the county staffs and contradicted the basic concept of involving the counties at all, versus ABAG's performing the whole study itself. Overall ABAG staff support generally turned out to be greater than was anticipated.

As if these organizational problems were not damaging enough, the whole program just happened to occur during the worst recorded drought in California history! There really was never sufficient rainstorms to monitor, and the ones that did occur were not as large as the average, much less large enough sometimes to bracket the average.

Despite these difficulties 55 storms were sampled (out of 95 that were planned or hoped for), and almost 600 samples were analyzed from those events and 25 baseflow samples were tested. Thirteen residential watersheds were monitored, ranging in size from 0.2 to 164 square miles. Seven agricultural and open space areas were sampled, as were three industrial sites and one commercial area. Most of the samples were taken in streams or storm drains at the lower ends of watersheds, so much integration of mixed and reacted constituents from streets, lawns, paved areas and other surfaces is reflected in the data collected.

The constituents analyzed varied somewhat, but the "core parameters" were 5-day BOD, total nitrogen, total phosphorus, suspended solids, volatile suspended solids, total dissolved solids, and lead. Other "spot-check parameters," sampled less frequently, included COD, fecal coliforms, fecal streptococci, total coliforms, cadmium, chromium, mercury, zinc, copper, and total identifiable chlorinated hydrocarbons (pesticides). Costs per sample for the core parameters ranged from $50 to $64.50, depending on which lab performed the tests. Total sample analyses ranged from $165 to $233 per sample.

Results were generally not startling. Five-day BOD's averaged 21 to 22 mg/l from the residential, commercial, and industrial (urban) areas, ranging from 7 to 40 mg/l; while BOD's from agricultural lands averaged only 8 mg/l and ranged from 4 to 13 mg/l. COD's were much higher, ranging from 14 to 195 mg/l, and the urban areas again had the higher means. Suspended solids were higher in the agricultural areas (786 mg/l) while the means for the urban land uses were 95, 158, and 252 mg/l for industrial, commercial, and residential lands, respectively. Total dissolved solids, on the other hand, were somewhat higher in the urban areas. The mean for residential areas was 176 mg/l with a range from 42 to 1906 mg/l. The mean for agricultural land was 109 mg/l with a range of 35 to 483 mg/l.

The nutrient levels (nitrogen and phosphorus) were found to be much higher from all land uses than was necessary for rapid growth of algae, but it was concluded (from data not presented) that the Bay still receives much more of these nutrients from point waste discharges.

Very large counts of bacteria occurred in the runoff from all areas, but the highest numbers were found in runoff from residential areas (110,000 per 100 ml-fecal coliforms), compared to 6,200 in agricultural runoff.

Concentrations of lead, zinc, and copper were quite high, ranging from 0.15 to 0.72 mg/l on the average. Almost 2 mg/l of lead were found in a sample of industrial area runoff, and 1.2 mg/l of zinc was found in a residential area. The other metals tended to be quite low, approaching the detection limits (0.005 to 0.01 mg/l).

With respect to direct satisfaction of the originally-stated objectives, each one was not fully attained. The mass emission rates were not satisfactorily estimated because a) there was too little rain; b) there was tremendous variability of concentrations, not only among watersheds, but at single sites; and c) the sampling program was discontinued after only one rainy season.

The model calibration (much less verification) was incomplete because not all the data were adequate for that purpose. The measured flows were often too small and the quality data were sometimes fragmented, incomplete, or inconclusive. Also not all of the sampled watersheds were modeled.

The local demonstration of the existence of water quality problems was left unresolved. While it was unarguable that concentrations of some constituents were high in the runoff from some watersheds, local decision makers fell back to the claim that adverse impacts from those concentrations had not been demonstrated. Indeed they had not. And that criterion (objective) had apparently never been raised before the monitoring began.

Lastly, the establishment of a continuing regional sampling network was not attained. No funding could be found to support that effort. On a subsequent but different EPA grant ABAG did collect 3 or 4 composite samples at four sites, but none of these were in the 1976-77 network.

In summary, the ABAG sampling program did provide certain data on urban runoff quality which had never been collected in the region previously, and these data provided a comparison of the area's runoff behavior with national average behavior monitored in other places. But the collected data did not lead to a clear consensus of whether runoff represents a significant water quality problem to be addressed by local governments.

THE NATIONWIDE URBAN RUNOFF PROGRAM

There is underway in 1980 a Nationwide Urban Runoff Program (NURP), sponsored by EPA and with major contributions from USGS. The program is fundamentally a follow-on to the nationwide 208 planning process for "nonpoint" pollution control evaluations. NURP's primary thrust is data collection in roughly 30 cities across the country. The data are to be used to a) gain fundamental understanding of the behavior of urban runoff quality, b) evaluate the effectiveness of various runoff management practices, and c) assess the impacts of urban runoff on the beneficial uses of various types of receiving waters.

Each of the 30 cities is defining its own specific objectives for the NURP project in its area. These depend on the problems generically identified in the past, the type of receiving water involved, and to some degree on the ambition of the local investigators. Some examples of objective statements* prepared for NURP studies in four cities follow, some of which admittedly were never undertaken:

1. "The project is aimed at determining what urban runoff controls can be used to prevent deterioration of lake water quality."

2. "The project proposes to test the efficiency of... catch basins and sand filters to ultimately revise the [existing] stormwater policy."

3. "The Great Lakes program is studying a number of BMP's in the area. Therefore NURP will concentrate on runoff characterization and source identification."

4. "The project...will assess the effects of storage detention on pollutant loads and will evaluate the potential load reductions due to source controls... The evaluation will lead to identifying the design factors necessary for siting, sizing, and operating storage facilities based on the analyses of runoff variables such as magnitude, duration, and recurrence of flow and pollutant loads."

There are, of course, some 26 other objective statements as well, ranging from complete holism and ambitious monitoring of air and water quality, cost and benefit evaluations, dry land and receiving water sampling, and evaluation of ordinances and legislation; to directed and simple statements of plans to monitor x number of catch basins and y months of street sweeping in z watersheds.

---

*Taken from unpublished memos provided by EPA

The NURP project here in the Denver area is one of the more ambitious ones. USGS is heavily involved in the work and is contributing funds to the study. Six to eight sites are being monitored for flow and water quality. Precipitation is to be mapped from radar-monitored storm tracks. BMP's will be implemented in some areas and resulting institutional problems will be analyzed. The South Platte River and a local lake will be monitored to develop estimates of the effects of urban runoff on receiving water. BMP's to be studied include detention basins and street sweeping. Special substudies planned include characterization of the relationships between total and soluble pollutant loads for several land uses, determinations of possible relationships between fractions of pollutant loads associated with discrete particle sizes of solids in runoff, and some studies involving a rainfall simulator. Hence, this study has undertaken to satisfy objectives ranging from pure research to design and construction of physical works. Moreover, transferability of our results to other areas is specifically to be illustrated, if possible. We remain hopeful, if not highly optimistic, as the study progresses. But it is trepidatious to be aware, as we are, that the monitoring planned and underway was not designed to accommodate the differences in sampling effort that these objectives imply.

In summary, the NURP program is relatively new. Indeed, it is brand new. Many of the projects are not yet [1980] fully underway, and all of them are scheduled to be over in a brief few years. Still it is encouraging that EPA and USGS have undertaken nationwide monitoring of urban runoff at this level of intensity. But the program will not provide all the answers that urban runoff investigators will ever need, and it should be clear that the program was not intended to do so. Local governments and land developers will still be expected in the future to perform the bulk of needed measurements. Local governments alone will likely provide the only long-term monitoring that will be done, and there are no long-term monitoring programs underway today.

EXISTING MONITORING GUIDELINES

The practice of water quality monitoring of urban runoff is by no means at ground zero, even if past experiences have been exploratory, short-term, and largely exercises in learning. There is a body of reports, papers, and books several feet thick on the topics of monitoring techniques, network design, statistical inferences, and laboratory analyses and quality control or quality assurance.

Probably the most comprehensive work on the subject is the "Areawide Assessment Procedures Manual" (U.S. EPA, 1976) published in July 1976 in three volumes at the height of the original 208 program. These volumes contain at least a thousand pages of technical guidance on all aspects of water quality management, largely but not entirely for nonpoint sources of pollution and with particular emphasis on urban stormwater. Topics included in this voluminous work include preliminary problem assessment, procedures for assessment of urban pollutant sources and loadings, evaluation of stream impacts, evaluation and selection of control alternatives, performance and cost data on various control methods, and a 325-page appendix on "monitoring requirements, methods, and costs."

Particularly good discussions in this appendix are given for measurement site selection criteria, general guidance on measurement frequency and sampling intervals, and sampling equipment features -- particularly advantageous ones. Guidance is not given on specifically how many samples should be taken for attainment of each objective, although specific objectives for monitoring are listed at the outset and described as essential for proper design of a monitoring program. Information on costs is primarily general and brief, though instructive; but specific cost guidelines are now out-of-date (which means they are low). One page is devoted to each of 168 individual quality constituents, describing generally why each one may be of interest in urban runoff (or anywhere); what quality criteria, if any, exist with respect to it; and what the sampling and lab analyses requirements are.

Another comprehensive report on methodology for the study of urban runoff quality and its control is the 1976 EPA report by Wullschleger, Zanoni, and Hansen (1976). This report, 326 pages in length, gives detailed guidance on sampling procedures, both for wet and dry periods; monitoring instruments; quality constituents and lab analyses; and methodology for evaluating storm discharges and various treatment methods. It was this report that was largely responsible for the impetus to undertake the work reported here. One of the recommendations of Wullschleger, Zanoni, and Hansen (1976) was:

> The recommended sampling programs [should] be made more specific as to number of samples, sample frequency, duration of sampling, etc., by means of a study based primarily upon a statistical approach.

A shorter, more generalized, but important and incisive set of guidelines for collecting, analyzing, and using urban stormwater data is that by Alley (1977). This guidebook was drafted and edited by Alley, but its value is enhanced by its having been reviewed, critically edited, and partially rewritten by 70 experienced urban water specialists during a week-long Engineering Foundation Research Conference in late 1976, convened especially for this purpose. As a result, it represents the best thinking, the state-of-the-art at that time (and in 1980 as well) on this topic. Chapters include separate discussions of data utilization (signigicantly, the first chapter), data analyses, network planning and design, instrumentation, and data collection. The chapter on network planning and design is especially germane to our considerations in this report.

The chapter is general in its guidance and never becomes specific about numbers of samples and the like; but the author is expansive on the need for setting explicit objectives first, recognizing what they are and when they change and how that affects the network, and even on a procedure for selecting the objectives. Perhaps most significant is the suggestion that monitoring objectives are statements of policy, not technique, and hence they should be established by policy makers and then satisfied by technicians.

This work throughout is also heavily directed toward monitoring of data whose intended analyses, regardless of the objective, would be through the use of mathematical models, calibration and verification of which become overriding objectives for the collection programs to be designed. Overall this booklet, only 115 pages long, is an excellent summary of all aspects of urban runoff sampling, even if it falls short of specific, numerical or statistical guidance on sample numbers and the like.

REFERENCES FOR CHAPTER III

1. Alley, William M., Guide for Collection, Analysis, and Use of Urban Stormwater Data, A Report on a Conference held at Tidewater Inn, Easton, Maryland, November 28 - December 3, 1976, American Society of Civil Engineers, New York, 1977, 115p.

2. Areawide Assessment Procedures Manual, EPA-600/9-76-014, Volumes I, II, and III, (3-ring binders), EPA Municipal Environmental Research Laboratory, Cincinnati, Ohio, July 1976.

3. Characklis, W.G., F.J. Gaudet, F.L. Roe, and P.B. Bedient, Maximum Utilization of Water Resources in a Planned Community: Executive Summary, EPA-600/2-79-050a, EPA Municipal Environmental Research Laboratory, Cincinnati, Ohio, July 1979, 136p.

4. Polls, Irwin, Instream Sampling Program, Research and Development Department, The Metropolitan Sanitary District of Greater Chicago, July 1978, 424p.

5. Polls, Irwin, Land Use Runoff Sampling Program, Research and Development Department, The Metropolitan Sanitary District of Greater Chicago, December 1978, 179p.

6. Polls, I., R. Lanyon, C. Lue-Hing, S.J. Sedita, and D.R. Zenz, "A Water Quality Sampling Program for 208 Planning in Northeastern Illinois," in Establishment of Water Quality Monitoring Programs, edited by L.G. Everett and K.D. Schmidt, AWRA, Minneapolis, 1978, pp. 328-338.

7. Polls, Irwin, and Richard Lanyon, "Pollutant Concentrations from Homogeneous Land Uses," Journal of the Environmental Engineering Division, ASCE, Vol. 106, No. EE1, February, 1980, pp. 69-80.

8. Wullschleger, R.E., A.E. Zanoni, and C.A. Hansen, Methodology for the Study of Urban Storm Generated Pollution and Control, EPA-600/2-76-145, EPA Municipal Environmental Research Laboratory, Cincinnati, Ohio, August 1976, 325p.

IV. THE STATISTICS OF MONITORING

SITES AND SAMPLES

As should be evident from all the foregoing discussions, the total number of samples required during the course of an urban runoff sampling program depends on:

1) the objective to be satisfied, and

2) a number of possible variables to be tested under each objective.

The number of samples to be taken (bottles to be filled with soil or water) each year can be computed for any objective from the following general expression:

$$\begin{aligned} n_a &= n_u \times n_w \times n_f \times n_e \times n_s + n_u \times n_w \times n_f \times n_d \times n_b \\ &= n_u n_w n_f (n_e n_s + n_d n_b) \qquad (1) \end{aligned}$$

in which

n_a = the total number of samples to be taken annually;

n_u = the number of separate urban areas, cities, to be sampled;

n_w = the number of watersheds or subareas to be sampled per urban area;

n_f = the number of sampling sites to be used in each watershed;

n_e = the number of rainfall-runoff events, storms, to be sampled each year;

n_s = the number of samples to be taken, bottles to be filled, at each site during a single storm;

n_d = the number of interevent, dry-period samples per sampling event; and

n_b = the number of dry-period sampling events per year.

The total number of samples taken for the entire sampling program over t years is

$$n = n_a \times t \qquad (2)$$

TABLE 1

SAMPLING PROGRAM CHARACTERISTICS FOR EACH OBJECTIIVE

Objective*	time, yr,t	#Urban areas,n_u	#Water-sheds, n_w	#Sites per w-shed, n_f	#Storms/ yr, n_e	#Samples/ storm, n_s	#Dry-period events/yr, n_b	#Dry-period samples/event, n_d
I	5	3	3	3	20	15	5	20
II	1	1	3	2	5	5	2	2
III	1	1	10	1	5	3	2	2
IV	1	1	1	1	2	10	2	2
V	25	1	10	1	20	3	0	0
VI	25	1	10	1	3	3	0	0

*I = Scientific research.
II = Definition of problems.
III = Analysis of alternative solutions
IV = Design of abatement measures.
V = Regulatory compliance monitoring.
VI = Operational performance monitoring.

TABLE 2

NUMBERS OF SAMPLING SORTIES REQUIRED FOR EACH OBJECTIVE

	Sorties/year*					Samples/	Total
Objective**	Type 1	Type 2	Type 3	Type 4	Total***	yr, n_a	Samples, n
I	90	45	270	270	675	10,800	54,000
II	8	4	15	15	42	174	174
III	13	7	25	25	70	190	190
IV	1	1	1	1	4	24	24
V	0	0	100	100	200	600	15,000
VI	0	0	15	15	30	90	2,250

*Type 1 = Dry period: 2-man crew for 8 hours = 16 man-hours.
Type 2 = Dry period: 4-man crew for 16 hours = 48 man-hours.
Type 3 = Storm event: 2-man crew for 24 hours = 48 man-hours.
Type 4 = Storm event: 4-man crew for 24 hours = 96 man-hours.

Assumption: 2/3 Type 1 + 1/3 Type 2 for dry periods;
1/2 Type 3 + 1/2 Type 4 for storm events.

**I = Scientific research.
II = Definition of problems.
III = Analysis of alternative solutions.
IV = Design of abatement measures.
V = Regulatory compliance monitoring.
IV = Operational performance monitoring.

***Sorties/yr = $n_u n_w n_f (n_e + n_b)$

TABLE 3

LABORATORY ANALYSES ASSUMED FOR EACH OBJECTIVE

Constituent	Objective* I	II	III	IV	V	VI
BOD or COD	x	x	x	x	x	x
Coliform organisms	x	x		x		
Dissolved oxygen	x			x		x
Nitrate-nitrogen	x	x		x		
Total organic nitrogen	x					
pH	x			x		
Phosphate, dissolved ortho-	x					
Dissolved solids	x					
Suspended solids	x	x	x	x	x	x
Total solids	x	x	x			
Total constituents, n_c	10	5	3	5	3	3
Laboratory analyses/yr, n_y	108,000	870	570	120	1,800	270
Total lab analyses, n_t	540,000	870	570	120	45,000	6,750

*I = Scientific research.
II = Definition of problems.
III = Analysis of alternative solutions.
IV = Design of abatement measures.
V = Regulatory compliance monitoring.
VI = Operational performance monitoring.

From the discussion of objectives in Chapter II and the discussions of successes and failures in Chapter III, minimal values of each variable in equations 1 and 2 are given in Table 1.

Several things should be noted about the values in this table. First and foremost, the numbers represent what the writers consider to be the minimum numbers of variables to satisfy each objective, with one exception. The exception is the 20 storms per year at a site under objective 1, which may be difficult to obtain in the semi-arid parts of the West. But perhaps that means merely that sampling in the West would have to persist for more years to satisfy this objective. Secondly, the numbers of samples and events signify the numbers of "good" events in which sampling is carried through to the end of the event. Every experienced sampler is aware that many times sampling sorties must be abandoned in mid-sampling, either because equipment fails, the rainfall or runoff proves to be insignificant, or any of the countless other things go wrong. ERG's experience has been that 50 percent more sorties than "good," completed ones can be anticipated, and 100 percent more is not rare, but merely indicative of a string of bad luck. Lastly, the numbers of years for sampling, t, are just that and do not include time for site selection or installation of any needed weirs or other equipment.

All sampling sorties to the field are not likely to be alike in duration and intensity. Dry-period accumulation sampling, for example, might be done with a single half day trip to each site, while a large storm event might require several crews alternately sampling around the clock for two full days. The results of the indicated multiplications to solve equation 1, divided into nominal but again good sampling sortie types, are shown in Table 2.

It becomes clear from this table that there are tremendous differences among these objectives where sampling is concerned. The range of required samples is 24 to 54,000. The sampling programs range in duration from 1 year (and perhaps less) to 25 years (and perhaps perpetuity). The field trips or sorties per year range from 4 to 675 (that's almost two per day, and this particular program would last for five years!)

QUALITY CONSTITUENTS

The numbers of field sorties and sample bottles are already quite large for some objectives, but the constituents to be measured in each bottle make the laboratory analyses numbers even bigger.

Table 3 summarizes the constituents that might be analyzed in a minimal program to satisfy each objective. Also shown is the total number of laboratory analyses required per year for each objective, which can be computed as

$$n_y = n_a \times n_c \qquad (3)$$

where

n_y = the number of laboratory analyses required annually, and

n_c = the number of constituents to be analyzed in each sample bottle.

The total number of lab analyses required over the sampling program duration can be computed as

$$n_t = n_y \times t \qquad (4)$$

The constituents listed in Table 3 clearly do not form an exhaustive list. There are no metals listed. Temperature is not even listed, although it can be assumed that water temperature would be measured and recorded routinely in the field at least once as part of each overall sortie. Exotic organics are not listed. But the ten constituents given are considered the minimum requirements that an investigator might set for himself under a scientific research objective, even though in Chapter II we said that 20 constituents could be required. [An analysis of that possibility is given in Chapter VI.]

ASSISTANCE FROM STATISTICS

Thus far in this chapter we have merely specified numbers of desirable measurement variables, such as watersheds, sampling sites, constituents, and the like, without supporting analysis. One of the major thrusts of this work has been that sampling program design should be predicated on the objective for sampling at all; and a corollary of that is that sampling program design should be based on the intended uses of the data, which means it should be based on the interpretive analyses to which one intends to subject the data once collected.

If interpretation of collected data is to be meaningful, fairly simple, and reliably successful, two steps should be taken before going to the field:

1) The objective should be restated in terms of hypotheses or assumptions that are to be tested. For example, it is difficult to know just what to do, where to go to sample, how many samples to take, and what to measure, if the objective is never stated more specifically than: "Is runoff from our city causing a problem?"

It is much simpler to visualize and to go about demonstrating one way or the other whether *this* problem statement is true or false: "Concern has arisen that runoff from our city is adversely impacting downstream beneficial uses of our receiving water. Among these is recreation and aesthetics. If the concern is well-founded then runoff is causing serious or repeated exceedances of quality criteria for recreation and aesthetics. These include the existence anytime of coliform organism counts of greater than 1000 MPN/100 ml, suspended solids greater than 50 mg/l, and visible, persistent, monthly summertime algal biomass concentrations greater than 1 mg/l or chlorophyll-a concentrations greater than 0.02 mg/l liter in unfiltered water. Higher summer-period loadings than those necessary to make nitrate-nitrogen concentrations above 0.3 mg/l and orthophosphate concentrations in excess of 0.01 mg/l could be taken as indicators of potentially excessive algal growth. We will now test the hypothesis that these concentrations are being exceeded as a result of our urban storm runoff." [Similar hypotheses could be stated and tested about effects on other beneficial uses.]

2) *To the extent possible, the results of the monitoring should be anticipated*. If, for example, in the hypothesis above, it were decided that nitrate-nitrogen and phosphorus should be measured in the stream instead of making the more difficult algal measurements, very reliably and repeatedly low nutrient concentrations might be found, but a wrong conclusion could be drawn. In hot summer periods at the slow-moving, "old swimming hole," algae may be blooming profusely. In so doing they will likely be extracting virtually all of the nitrogen and phosphorus available. So low concentrations in the water, well below the rule-of-thumb criteria, may result. Still the nitrogen or phosphorus loads from upstream, such as storm runoff, may be significantly high. So, *that* is the place to measure nitrogen and phosphorus. Additionally, some analytical measurements are much less precise than others, e.g., the variability of samples of coliform organisms is likely to be much broader than the variability in dissolved solids measurements, even among test samples from the same water or even from the same sample bottle. So decisions about sample numbers should be based in part on the constituents having the least expected measurement accuracy. Finally, if the objective is amenable to algebraic expression, such as a chemical equilibrium equation or even a simple mass balance or an anion/cation balance, it should be so stated. A mathematical model to be calibrated is its own expression of the calibration objective. Such expressions lead the

sampler directly to the pigeon-holes of variables that must be filled by sampling. Wasteful sampling of all but one or two critical variables can be avoided, merely by stating the objective, the hypothesis, the problem in algebraic terms first so all the important variables are anticipated.

Statistics of Sample Size

As we are about to see, statistical theory of sampling is quite limited in its ability to guide decisions about sampling to satisfy the six objectives being pursued here. But statistical tests and inferences should not be ignored, and indeed they are helpful for judging the utility of some measured records.

The most fundamental question we might want to subject to statistical theory is, how many samples must be taken to assure a reliable, believable result from the measured data?

This question cannot be answered without some measured data! Astonishingly, we have to know (or project or guess or anticipate) some things about the data we are going to measure before we can know that we have measured enough. To appreciate why that is, consider these definitions.

The average or mean of a set of measurements is

$$\bar{x} = [\sum_{i=1}^{n} x_i]/n \tag{5}$$

in which

$\bar{x}$ = the average value of the measurements made;

x_i = the value of each (the i-th) measurement; and

n = the number of measurements.

The sample variance is defined as

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{6}$$

The value of s itself, the square-root of the sample variance, is the standard deviation of the sampled set of measurements, and it describes the degree to which a random set of values measured for essentially the same property, such as the pH or COD of storm runoff, will vary from the average value of those same measurements.

A similar and related indicator of the expectable scatter among measurements is the *coefficient of variation*. It is estimated as

$$c.v. = \frac{s}{\bar{x}} \qquad (7)$$

The mean, the variance, and the standard deviation of a set of sampled measurements can be appreciated to be *estimates* of the same values for all possible measurements of the same property, which collectively are called, the *population*. It is the *population* of pH or COD or nitrate values that a finite, even small, set of measurements is meant to *sample* and hence estimate. Herein the population parameters will be designated with capital letters. Therefore

$\bar{X}$ = the population mean;

S^2 = the population variance; and

S = the population standard deviation.

The Student's t Test

There is available generally in statistics a "standardized" (or normalized, dimensionless) distribution called the "Student's t" distribution (see the values given in Appendix I). The values of t given in the appendix are the normalized ratios for any set of samples of any property as defined by:

$$t = \frac{\bar{X}-\bar{x}}{s}\sqrt{n-1} = \frac{\bar{X}-\bar{x}}{S}\sqrt{n} \qquad (8)$$

Recall that we are seeking the value of n. Equation 8 can obviously be rearranged to give an algebraic expression for that elusive quantity, as

$$n = \frac{t^2s^2}{\bar{X}-\bar{x}} + 1 = \frac{t^2S^2}{\bar{X}-\bar{x}} \qquad (9)$$

but that expression contains the mean of the samples we have not yet taken, the variance of those samples, and the population mean, which is something we would love to estimate, if we could just figure out how many samples to take!

Well, as confusing and frustrating as this appears, all is not completely hopeless. The Student's t test will be helpful. But already it is evident that some trial sampling or literature-reported values from other studies, plus some estimates of variability will also be necessary, or n cannot be computed.

Futhermore, as reference to the values of t in Appendix I will suggest, the values of t are a function of another variable called "degrees of freedom."

Although defining degrees of freedom in multiple effect tests becomes involved, in general the degrees of freedom are the number of observations, n, minus the number of parameter values to be estimated. Often we are testing for the value of the mean only, and in that case the degrees of freedom involved are merely n-1.

Still further, t is defined, as the appendix shows, in part by a "confidence interval," which must be specified. For example, one might want to design an experiment to be 95 percent "confident" that the population mean is adequately estimated by the mean of only the sampled values. This is the same as saying that one wants only a 5 percent chance that the sample mean is not an adequate estimation of the population mean. Since there is a 2.5 percent chance that the sample mean would be too high and a 2.5 percent chance that the sample mean would be too low, the confidence level column in the appendix to be used is defined by

$$t_c = (100-a/2)/100 \qquad (10)$$

where the subscript c denotes the confidence interval (95), t_c is the column to be used to find the appropriate value, and a is the percentage error permitted (5). Therefore to test the 95 percent confidence interval, values of t are actually selected from the column denoted: $t_{0.975}$. Another way of looking at this is to consider the confidence interval as a definition of confidence limits on the population mean:

$$\bar{x} - \frac{t_c s}{\sqrt{n-1}} \leq \bar{X} \leq \bar{x} + \frac{t_c s}{\sqrt{n-1}} \qquad (11)$$

Here we can see that half the error is expected to be above the mean, and half is expected to be below it. Some examples will demonstrate one use of the Student's t test. Suppose we had a runoff sample set of flow-weighted BOD concentrations from 25 storms. Suppose further that the sample variance, s^2, is 100. We can estimate the population mean concentration at the 95 and 50 percent confidence levels with the Student's t test. With 24 degrees of freedom (n-1), the t values from Appendix I are: $t_{95} = 2.06$ (from the $t_{.975}$ column); and $t_{50} = 0.685$ (taken from the $t_{.75}$ column). The confidence intervals therefore are:

$$\frac{t_{95}s}{\sqrt{n-1}} = \frac{(2.06)(10)}{\sqrt{24}} = \pm 4.2$$

$$\frac{t_{50}s}{\sqrt{n-1}} = \frac{0.685(10)}{\sqrt{24}} = \pm 1.40$$

Therefore, if the sample mean for all 25 storms were 15 mg/l, we can say that it is 95 percent certain that the population mean (the mean BOD concentration for all storms) is between 10.8 and 19.2 mg/l. The probability is 50 percent, however, that the mean lies between 13.6 and 16.4 mg/l.

Now suppose we only had data from the first 10 storms. The mean might be still near 15 mg/l, say 16 mg/l; and the variance could well be larger, say 144. Then the confidence intervals would be

$$\frac{t_{95}s}{\sqrt{n-1}} = \frac{(2.06)(12)}{\sqrt{9}} = \pm 8.24$$

$$\frac{t_{50}s}{\sqrt{n-1}} = \frac{0.685(12)}{\sqrt{9}} = \pm 2.74$$

and we can see that a smaller sample size has led to a much broader range of expected values of the mean value of the entire population. We can now be 95 percent assured that the population mean lies somewhere between 7.76 mg/l and 24.24 mg/l; and we can be 50 percent assured that the true mean lies between 13.26 mg/l and 18.74 mg/l. At the same levels of confidence, then, our degree of accuracy has diminished by having a smaller sample size.

Therefore, we can learn something from the Student's t test about the number of samples necessary to be collected to yield a certain confidence about the results obtained. But we see that to make these computations and comparisons we must have some data on hand initially to give a mean value for the sample, a sample variance, and a standard deviation. This largely begs our initial and fundamental question, which was: How many samples do I need to collect?

The Student's t test can be massaged slightly to give, perhaps, some more useful insights. If we recall that the coefficient of variation is $s/\bar{x}$, and if we define a fractional percentage error of the true mean that we will accept as

$$P = \frac{\bar{x}-\bar{X}}{\bar{x}} \quad ,$$

then we can redefine t from Equation 8 as

$$t = \frac{\bar{x} - \bar{X}}{(s^2/n)^{1/2}} = \frac{\dfrac{\bar{x} - \bar{X}}{\bar{x}}\sqrt{n}}{s/\bar{x}} = \frac{P\sqrt{n}}{c.v.} \qquad (12)$$

We also assume in this rearrangement that the sample variance, s^2, is a reasonably good approximation of the true population variance, S^2, which is difficult to discern.

But now the number of samples can be defined as follows, merely by rearrangement of equation 12.

$$n = \frac{t^2 \ c.v.^2}{P^2} \qquad (13)$$

and we can make some reasonable assumptions about c.v. and P and check for various values of n and t in the Appendix.

For example, we could say we would like to sample BOD's such that the standard deviation would be not greater than the mean value, i.e., c.v. = 1. Also we could reasonably expect the sample mean to be within 20 percent of the true mean value. Indeed we could say we wish to be 95 percent confident that this is so. A mid-range value of t can be selected (note that they vary only slightly at a given confidence level above 3 or 4 degrees of freedom), or even a low value near the upper range of degrees of freedom could be selected, wherein the values are close to those of a normal distribution, and a rough estimate of n can be made. For this example, taking $t_{95} = 2.0$,

$$n = \frac{(2)^2 \ (1)^2}{(0.2)^2} = 100 \text{ samples}$$

Now measurements of various constituents will show that the coefficient of variation varies over a broad range. Sherwani and Moreau (1975) have shown that bacterial sample counts are so variable that the coefficient of variation could be as high as 3, while pH is such a stable property that its c.v. could be as low as 0.05. For BOD they found c.v. to be 0.5 for their samples (which would lower the estimate in our example to 25 samples). Suspended solids had a c.v. of about 1.5. The value of c.v. for nutrients such as nitrogen and phosphorus was less than 1.0, actually about 0.6 to 0.7.

So this technique will lead to different indications of sample numbers required for each constituent or property (including flow). Sample (bottle) numbers should be planned according to the constituents with the higher coefficients of variation. If costs for lab analyses become prohibitively high, those constituents with lower c.v. values could be analyzed less frequently with the same degree of confidence maintained.

OTHER STATISTICAL TESTS

There are numerous statistical tests available for aiding judgments about the suitability, sufficiency, adequacy, and accuracy of measured data; and statistical techniques can be married with sampling design before measurements are made. These techniques quickly become involved, both in jargon and algebra, and little would be served by introducing the details here. A number of textbooks on the subjects are available. Two that have proved helpful to us are a statistics book geared to hydrologic data in particular by Haan (1977), and a text on experimental design by Hicks (1973).

Moreover, detailed introductions of statistical tests are presented herein in an Appendix prepared by CDM, which is directly related to problems encountered with our six objectives concerning urban runoff quality sampling. (See Appendix IV.)

At this point, however, several standard, well-known, and useful tests are introduced to provide the general reader an appreciation for the variety of ways in which statistical tests and inferences can be helpful.

The F Test -- For Testing the Equality of Variances

It often occurs that one wishes to compare the variability of samples from two (or more) sets of samples. For example, we might wish to compare the variability of measured suspended solids data from one watershed with those from another watershed; or we might want to compare the variability of solids concentrations on the rising limb of a storm's hydrograph with the variance of those samples collected during the recession, after the peak.

Population variances may be tested for statistically inferred equality with the "F test" or by using the "F distribution." As with the t test, standard tables of F-values are generally available. A set of these tables is given in Appendix II.

In this test, the sample statistic, F, is defined as the ratio of the two sample-set variances:

$$F = \frac{s_1^2}{s_2^2}$$

where the subscript 1 denotes the set having the larger variance, and 2 denotes the sample set with the smaller variance.

Each sample set, denoted by the subscripts, 1 and 2, can have different degrees of freedom which must be known. If these are $v_1 = n_1-1$, and $v_2 = n_2-1$, where n_1 and n_2 are the numbers of samples in the two sets, then the table of F values should be consulted at the row and column corresponding to $F[(1-a),v_1,v_2]$ and $F[a,v_1,v_2]$ with v_1 representing the "numerator" degrees of freedom and v_2 being the "denominator" degrees of freedom. The value of a, as before, is the decimal fraction of the acceptance level, say 5 percent or 0.05, and 1-a is the confidence level, say 95 percent or 0.95.

If the computed value of F is greater than the value found in the table, it can be reasonably concluded that the variances are not equal; or, as statisticians say, the hypothesis that they are equal may be rejected.

To find the correct value of F for the lower tail of the distribution, $F[a,v_1,v_2]$, the appendix must be read at the level of confidence with the numerator and denominator degrees of freedom reversed; and the reciprocal of the F value found is the correct value. That is

$$F[a,v_1,v_2] = \frac{1}{F[1-a,v_2,v_1]}$$

This procedure must be noted, since only values of F for confidence levels equal to or greater than 90 percent are given in the appendix.

Some examples should clarify the procedure. First, suppose that we wish to compare the variances of suspended solids samples from a single storm for equality between the samples taken before the peak of the runoff and those taken after the peak. If there are 8 samples taken before the storm peak and 6 values taken after the peak, and it is presumed that first flush would lead to a higher variance in the early period, we can find the two values of F from the appendix at, say, the 90 percent confidence level which will bound the acceptance region for the field-determined value of $F[s_1^2/s_2^2]$. From the table in the appendix,

$$F[0.95,8,6] = 4.15$$

$$F[0.05,8,6] = \frac{1}{F[0.95,6,8]} = \frac{1}{3.58} = 0.279$$

Therefore if the computed F value is less than 4.15 and greater than 0.279, we can conclude with 90 percent confidence that the variances are equal, statistically speaking. If the value is greater than 4.15 or less than 0.279 we are likewise 90 percent confident that the two variances are not equal. Notice that we are not at liberty to say in that case, however, that we are 90 percent confident that first flush occurred. Statistics alone tells us nothing about cause and effect for variances being equal or not equal.

Suppose also that we have 14 samples from each of two storms and we wish to compare the suspended solids loads from each one. Suppose the mean concentrations are similar, say 151 and 149 mg/l; but the variances are 52 and 139 respectively. We can compare the means with the t test and compare for equality of the variances with the F test. We shall use the 95 percent confidence level in both cases.

First we define a "pooled variance" of the two samples as

$$s_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

The t statistic for testing the equality of the two means is then

$$\frac{\bar{x}_1 - \bar{x}_2}{s_w\sqrt{1/n_1 + 1/n_2}} < t_{1-a,n_1+n_2-2}$$

where the t value is that in Appendix I in the column for 1-a/2, or 0.975 in this case. Therefore

$$s_w^2 = \frac{(14-1)(139) + (14-1)(52)}{14+14-2} = \frac{2483}{26} = 95.5$$

$s_w = 9.77$

$$\frac{\bar{x}_1 - \bar{x}_2}{s_w \sqrt{1/n_1 + 1/n_2}} = \frac{151 - 149}{9.77 \sqrt{1/14 + 1/14}} = 0.542$$

$t_{0.95,26} = 2.06$

Since the computed t value is less than the tabulated value we can accept the hypothesis that the two means are statistically identical, with 95 percent assurance.

Now the F test can be used to compare the variances.

$$F = \frac{s_1^2}{s_2^2} = \frac{139}{52} = 2.67$$

From Appendix 2

$F[0.95,14,14] = 2.9$ (by interpolation)

$F[0.05,14,14] = 1/2.9 = 0.34$

Since the computed value of F falls within the range between 0.34 and 2.9, we can also conclude with 95 percent confidence that the variances are statistically equivalent. That is, we can be fairly confident that the two storms are from among the same population of storms of this type.

The Chi-Square Test -- For Testing Sample vs Stated Variances

Another test that is often desired is the determination of whether or not a population variance is equal or not equal to a known or assumed or stated variance. This hypothesis can be tested with use of the chi-square distribution.

The chi-square test is based on the fact that the ratio of a sample variance and the population variance, multiplied by the sample degrees of freedom, has a predictable distribution. That is, the quantity

$$\frac{(n-1)s^2}{S^2}$$

has a chi-square distribution with n-1 degrees of freedom.

With a given value of a population variance, S_o^2, a test statistic may be computed; and, analogous to the t-distribution, tabulated values of chi-square may be read from Appendix III as the bracketing values. That is, the test statistic will be

$$x_c^2 = \frac{(n-1)s^2}{S_o^2}$$

and the hypothesis of equality of S and S_o is accepted if

$$x^2_{a/2,n-1} < x_c^2 < x^2_{1-a/2,n-1}$$

With the same distribution one can compute the confidence limits on the population variance:

$$(n-1)s^2/\, x^2_{1-a/2,n-1} < S^2 < (n-1)s^2/x^2_{a/2,n-1} \qquad (14)$$

Note that the values of chi-square for the confidence level, 100(1-a), say 95 percent, are read from the columns in the appendix headed by 1-a/2 (0.975) and a/2 (0.025).

Here is an example. Suppose we wish to sample a number of storms at a single point and that we wish to sample enough to be 95 percent confident that the variance in average COD values over each storm is 300, which we have read is about the variance in COD values found in other places across the nation. From samples from 25 storms we compute that the sample variance is 319. Is that sample variance equivalent statistically to the variance reported by the other investigators? Since it is numerically larger, can the 25 samples possibly be sufficient?

First we will state the hypothesis that the two variances are equal, $S=S_o$, and we shall test that hypothesis with the chi-square test.

$$x_c^2 = \frac{24(319)}{300} = 25.5$$

From Appendix III

$$\chi^2_{0.975,24} = 39.4 \quad \text{(by interpolation)}$$

$$\chi^2_{0.025,24} = 12.4$$

Because the value of χ_c^2 falls between 12.4 and 39.4, we accept the hypothesis that, with 95 percent confidence, $S=S_o$.

Moreover, we can compute that the true population variance lies within a range (and hence is not actually equivalent to the value of 300 that we found in the literature):

$$\frac{24(319)}{39.4} < S^2 < \frac{24(319)}{12.4} \qquad (15)$$

Thus the population variance lies between 194 and 617, again with 95 percent assurance.

There are a number of interesting things to note from the chi-square distribution and its use.

First it is worth noting that the range of acceptable values for the population variance is quite large. In the example the range varies by a factor of 39.4/12.4 = 3.18. It is also interesting that the value of the stated variance (300) does not appear in equation 15, and the result is that the upper and lower limits are not symmetrical about the stated population variance (617-300 = 317; 300-194 = 106). Nor for that matter are they symmetrical about the value of s^2, which does appear in equation 15 (617-319 = 298; 319-194 = 125).

Furthermore it is instructive to note that the test statistic will be equal to the degrees of freedom if the sample variance is actually numerically equivalent to the stated population variance. That is

$$\chi_c^2 = \frac{(n-1)\, S_o^2}{S_o^2} = n-1; \quad s^2 = S_o^2$$

In the example, chi-square = 25.5, which was very nearly equal to the degrees of freedom (24), because the two variances being tested for equality were numerically close to equivalent (300 vs 319). Thus prior to any sampling, one could consult the chi-square table in Appendix III and gain some insight to the numbers of samples required for the sampled population variance to be (roughly) equivalent to a stated overall population variance.

It can be seen that $\chi^2_{.50}$ is roughly equivalent to, but slightly less than, the degrees of freedom throughout the range of degrees of freedom. The values of $\chi^2_{.90}$ are slightly larger than the degrees of freedom throughout the range, with the value being 2.71 times as large as the degrees of freedom at n-1 = 1 and 1.2 times as large at n-1 = 100. As the degree of confidence is increased toward 99 percent, the χ^2 value becomes larger than the degrees of freedom by a greater amount. Correspondingly, the χ^2 value at the other end of the distribution, toward $\chi^2_{0.01}$, becomes smaller. This illustrates that as the stated confidence level becomes higher, the range of acceptable values becomes larger too. It is to be appreciated, therefore, that statistical significance at the 1 percent level is less stringent or demanding than at the 5 percent level and not the other way around as one might expect.

For the purpose of guidance, one can compute the values of $n-1/\chi^2$ for a number of confidence levels and degrees of freedom and seek those combinations where the ratios are greater than, let's say, 0.5 for the lower limit in equation 14 and less than 2.0 for the upper limit. Such computations are shown in Table 4. It can be seen that, from among the combinations given, these arbitrary criteria are met first at n-1 = 24 (or n = 25) at the 95 percent confidence level, and at n-1 = 12 (n = 13) for the 80 percent level. The range of values narrows toward 1.0 as the level of confidence decreases and the degrees of freedom increase. At 80 percent confidence or above the ratio never closely approaches 1.0. It is interesting to note that at the 95 percent confidence level something over 50 samples are required to reduce the upper limit ratio to 1.5.

Other insights to sample numbers and confidence levels to be gained from the chi-square distribution, and the t-distribution as well, are given in Appendix IV.

TABLE 4

RATIOS OF DEGREES OF FREEDOM TO CHI-SQUARE WHICH APPEAR IN THE COMPUTATION OF CONFIDENCE LIMITS ON VARIANCE

(Values of $n-1/X^2$)*

	1-a/2	n-1 Degrees of Freedom				
100(1-a)	a/2	100	50	24	12	8
98	0.99 0.01	0.74 1.43	0.66 1.68	0.56 2.20	0.46 3.36	0.40 4.85
95	0.975 0.025	0.77 1.35	0.70 1.54	0.61 1.94	0.52 2.72	0.46 3.67
90	0.95 0.05	0.80 1.28	0.74 1.44	0.66 1.79	0.57 2.29	0.52 2.93
80	0.90 0.10	0.84 1.21	0.79 1.33	0.72 1.53	0.65 1.90	0.60 2.29

* Paired values times sample variance yield upper and lower limits on population variance. Note range is narrowed by moving downward and to the left in the table.

REFERENCES FOR CHAPTER IV

1. Haan, Charles T., Statistical Methods in Hydrology, The Iowa State University Press, Ames, 1977, 378p.

2. Hicks, Charles R., Fundamental Concepts in the Design of Experiments, Second Edition, Holt, Rinehart and Winston, New York, 1973, 349p.

3. Sherwani, J.K., and D.H. Moreau, Strategies for Water Quality Monitoring, Report No. 107, Water Resources Research Institute of the University of North Carolina, UNC-WRRI-75-107, Raleigh, June 1975, 106p.

V. THE COSTS OF MONITORING

SAMPLING IN THE FIELD

Costs for sampling, laboratory analyses, and interpretation of results have been estimated based on recent experiences of ERG. Their data are not reproduced here verbatim as current price lists, but the unit costs reported have been generalized from their actual experience.

Table 5 shows the anticipated costs (in 1980 dollars) for the field sampling work required for each objective, which was shown in Chapter IV in Table 2. The values in the table are based on assumptions of 20 miles of travel per sampling sortie, 35 cents per mile for auto-van-truck expenses, and 28 dollars per man-hour as an average labor cost for field sampling crew members. The sampling man-hours were derived from the four types of sorties postulated in Table 2.

The results of this analysis show that field monitoring costs could vary from about $5,800 per year for objective IV (design of abatement measures) to over a million dollars per year for objective I (scientific research).

LABORATORY ANALYSES

There are hundreds, even thousands, of laboratory tests for chemical and physical properties or constituents of water that are available for analysis through commercial laboratories. The ranges of prices for these are quite broad, depending on how difficult or time-consuming each test is, the sophistication [expense] of the laboratory equipment required to perform each test, and how many samples are to be analyzed at a time. The economies of scale to be realized in a single laboratory by providing large numbers of samples on a continuing, predictable schedule are sizeable. They are far larger (50 to 80 percent savings over small sample sets) than one could expect to find in shopping for a less expensive laboratory. The business is rather highly competitive, and most savings are likely to be realized by 1) having large sample sets, and 2) using a laboratory whose field crew is near the site or sites to be sampled (if the lab's crews are to be used).

Table 6 lists generalized cost data for twenty constituents likely to be analyzed in storm runoff. Costs are shown for analyses of both wet and dry samples, and the effects of large sample sizes on unit costs are made evident. The number, 500, was chosen arbitrarily. A given lab will define a "large" sample set for which discounts will be offered according to its own workload, the total size of the job, and its own other proclivities. But unit savings for large sample sizes are routinely

TABLE 5

FIELD SAMPLING EFFORT AND COST FOR EACH OBJECTIVE

	Sampling Effort/Year			
Objective*	Miles**	Man-hours	Equivalent Man-years	Cost/yr, $***
I	13,500	42,480	20.0	1,194,165
II	840	2,480	1.2	69,734
III	1,400	4,144	2.0	116,522
IV	80	208	0.1	5,852
V	4,000	14,400	6.9	404,600
VI	600	2,160	1.0	60,690

*Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

**Assumption: 20 miles per sampling sortie; wet or dry period.

***$0.35 per mile + $28.00 per man-hour for field crews; 1980 data.

TABLE 6

SELECTED LABORATORY ANALYSES COSTS
1980 DOLLARS

	Unit Cost Per Liquid Sample,$		Unit Cost Per Solid Sample,$	
Constituent	Up to 500 Samples	More Than 500 Samples	Up to 500 Samples	More Than 500 Samples
COD or BOD	$ 12.50	$ 4.00	$ 22.50	$ 8.00
Coliforms	14.00	3.00	24.00	6.00
Lead	8.00	4.00	18.00	8.00
Nitrogen, ammonia	10.00	5.00	20.00	10.00
Nitrogen, nitrate	16.50	6.00	26.50	10.00
Nitrogen, total org.	18.00	5.00	28.00	10.00
Dissolved oxygen	6.00	2.00	16.00	4.00
pH	2.50	0	12.50	0
Phosphorus, total	11.00	5.00	21.00	10.00
Phosphate, ortho-	11.00	3.00	21.00	6.00
Solids, dissolved	6.50	3.00	13.00	6.00
Solids, suspended	6.00	3.00	12.00	6.00
Solids, total	6.50	4.00	13.00	8.00
Zinc	8.00	4.00	16.00	8.00
Subtotal	$136.50	$ 51.00	$263.50	$100.00
Pesticides				
PCB's Toxaphene Chlordane DDT's	120.00	80.00	240.00	160.00
Aldrin Heptachlor	40.00	30.00	80.00	60.00
Total, all constituents	$296.50	$161.00	$583.50	$320.00

available.

Table 6 shows a number of notable things. First, lab analyses are expensive. To analyze 50 bottles for all twenty constituents would cost almost $15,000. To realize the economy of scale, by showing up with 500 samples to have analyzed for all 20 constituents would cost $78,000. Those costs are for liquid samples. Dry-period, solid samples would cost twice as much.

Furthermore, the costs vary tremendously according to which constituents are analyzed. Total organic nitrogen costs more than seven times the analysis for pH. Four pesticides cost roughly the same as the 14 less exotic organic and the inorganic properties. So what-to-measure is worthy of consideration both from a scientific but also practical financial viewpoints.

One potentially important group of constituents not shown are the metal ions, such as chromium, iron, and mercury. The costs for analyses of these three metals, in 1980 dollars, are roughly 15, 8, and 18 dollars, respectively. Other possible constituents of interest might be total organic carbon ($20), hardness ($6), oil and grease ($13), and sulfate ($10). The prices in parentheses are for liquid samples in small numbers.

DATA INTERPRETATION

The labor costs for data interpretation will vary depending on who performs the interpreting. If the city has the qualified people on staff, the city may do its own analyses. If a consulting engineer or university professor is hired, the costs will likely be different, since the engineer's overhead and profit must be included. (The city has overhead costs as well, which may mean the costs will not be much different if this route is chosen.) Also, the commercial laboratory used to collect and analyze the samples may offer this service, and the advantages of having interpretations of data made by the same people who collected and analyzed the samples may be well worth the expense. Generally speaking, commercial lab unit rates are likely to be somewhat higher than city personnel rates, and consulting engineer rates are likely to be higher than commercial scientist rates. But the times available and required by these three sources of interpretation to perform the same analyses could result in the total costs for interpretation being roughly equal. This is a highly individualistic matter and should be analyzed on a case by case basis. It should be analyzed, though, before the field program of sampling is begun.

In this report, interpretation costs are estimated from, and generalized from, commercial lab rates derived from the past several years of experience at ERG. Interpretations are made at the commercial lab level by three kinds of staff members whose time is charged at different rates. "Laboratory staff" personnel are charged at a rate around $33.50 per hour, in 1980 dollars. Current charges for a "research scientist" are roughly $41.50 per hour. A "senior scientist," who would likely oversee all of the work but who would devote fewer hours than laboratory staff and research scientists, has a current hourly rate of $53 per hour.

We have assumed that an interpretation program would involve one-half day of senior scientist time for every 4 field sorties. Research scientist effort is estimated to be one-half day for each field sortie. Laboratory staff effort is anticipated to be 1 work day (8 hours) for every 4 sorties wherein the number of constituents to be analyzed is less than 6 (objectives II through VI). For samples requiring more than 6 constituents to be analyzed (objective I), laboratory staff time is estimated to be 1 work day for every 2 field sorties. These assumptions are just that, and the levels of effort for interpretation should be predicted on a case by case basis before the field program is begun.

INFLATION

The costs presented were derived from recent year's experiences (1976 through 1980). The costs have increased over that period, and they will surely rise again. Inflation is the main reason for anticipating higher costs, both for labor and for unit rates of per-constituent lab analyses. There are some analyses costs that actually reduce with time on a unit basis, because new equipment technology and automation can lead to favorable economies of scale. But a new piece of equipment that can perform 60 analyses per hour, compared to an older machine that could do 20, may cost 2.5 times as much, so the savings may be illusory. Savings to be realized in lab analyses will likely be from the invention of automatic equipment that will perform what was before a highly labor-intensive test. But the new equipment will probably be expensive, and its costs must be recovered. So these savings will not be gigantic. The largest savings to be realized will probably continue to result from discounts on unit lab rates that commercial laboratories will offer for analyses of large numbers of samples. These options should be explored in detail prior to establishing the sample numbers and constituents to be analyzed. An example of these effects is given in Chapter VI.

The recent results of inflation on lab tests, which have been reduced to straight-line estimators, are shown in Figure 1. Labor rate inflation for various levels of personnel are shown in Figure 2.

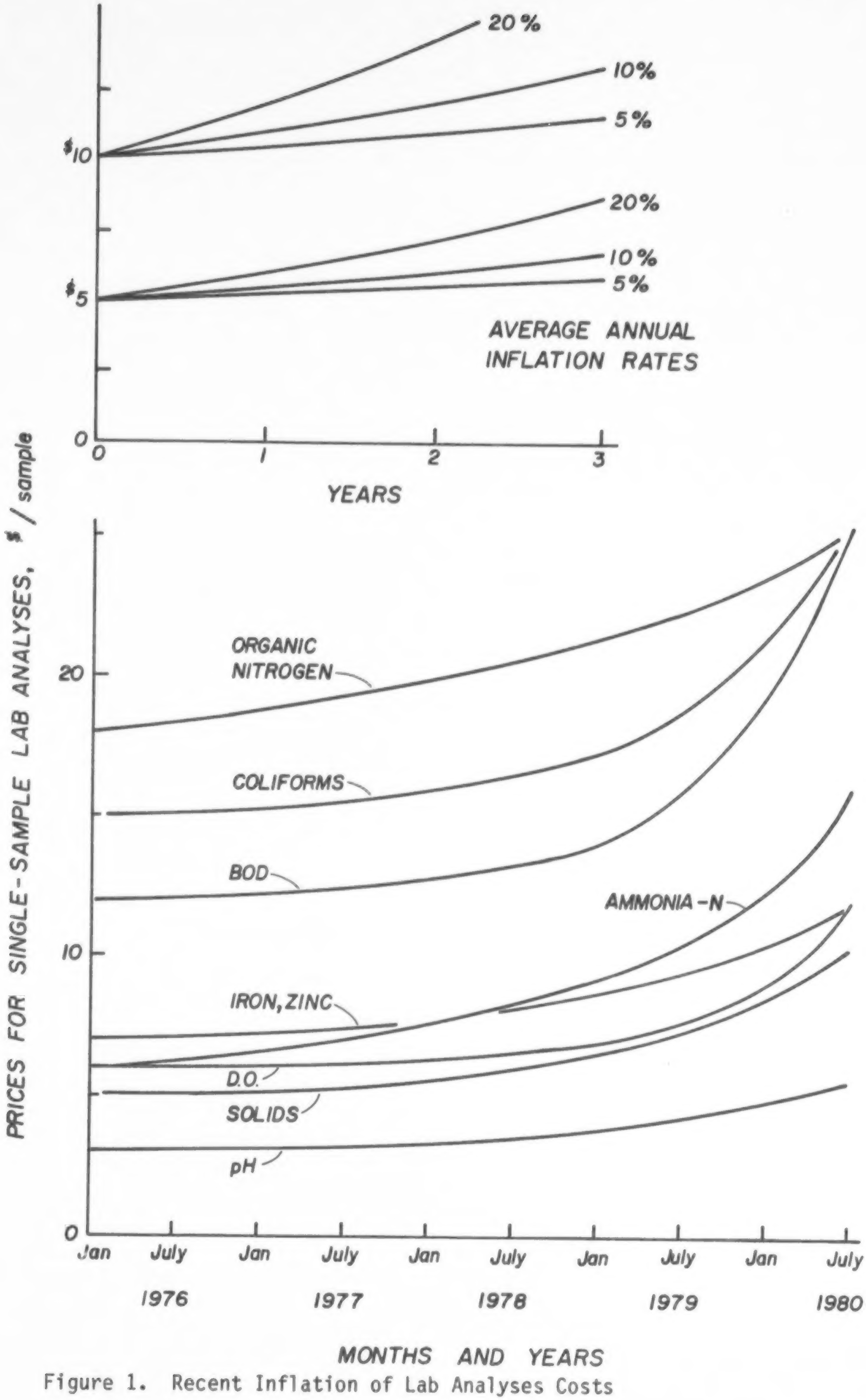

Figure 1. Recent Inflation of Lab Analyses Costs

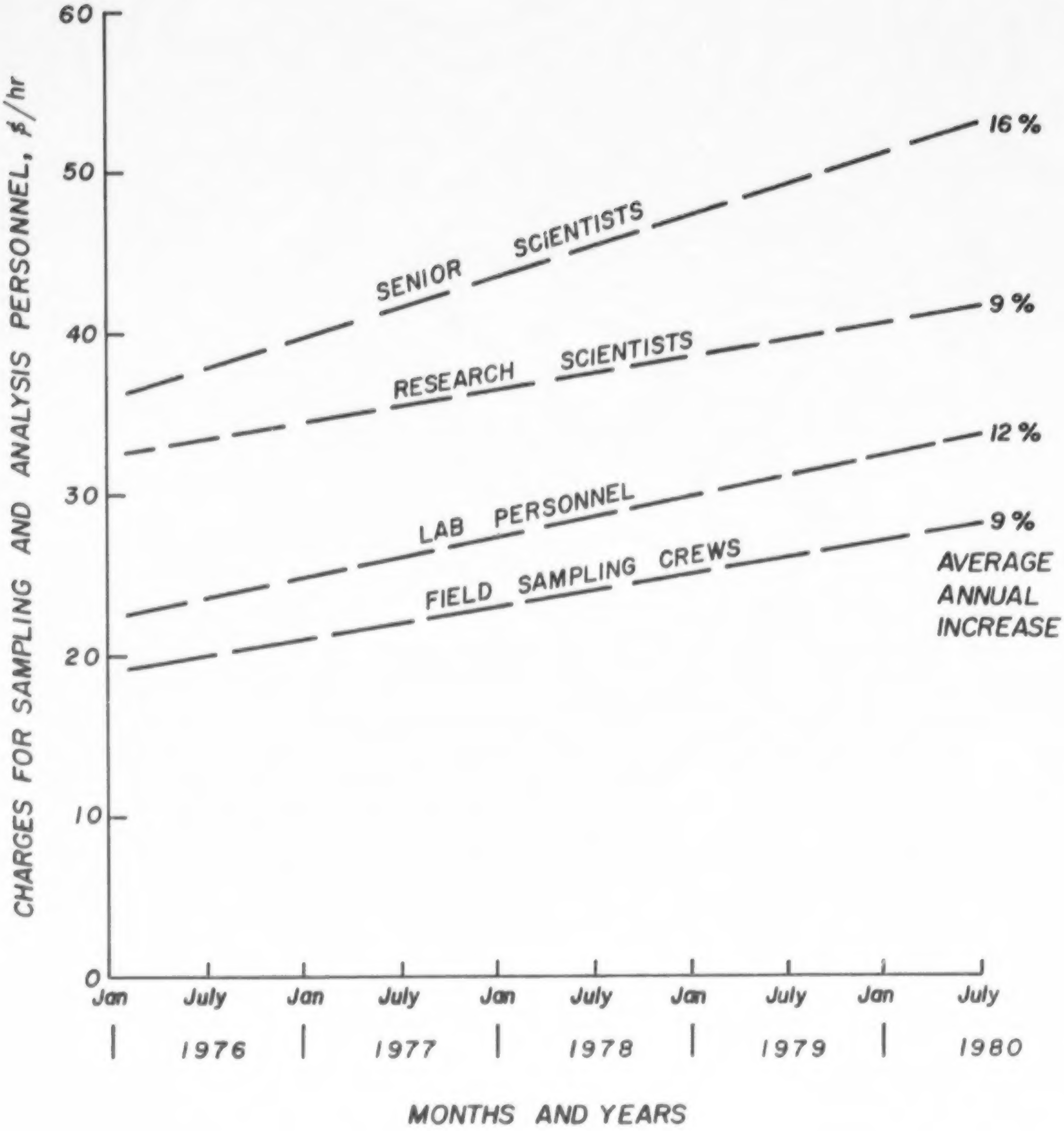

Figure 2. Recent Inflation of Data Interpretation Costs

VI. EXAMPLE PROGRAM DESIGNS

ANALYSES AND SAMPLES

Table 7 summarizes all the basic data for the six objectives described thus far. The total number of samples, as we have seen before, varies from 24 to 54,000. The total number of laboratory analyses for various constituents ranges from 120 to over half a million.

Table 8 lists the laboratory costs for each sample set, which assumes that the sample sets will each contain less than 500 samples. For each objective, then, the lab costs per sample would range from $25.50 for three fairly inexpensive constituents to over $100 for all ten constituents.

Table 9 summarizes the interpretation effort and cost for each objective.

COMPARATIVE COSTS OF COMPLETE PROGRAMS

Table 10 collects the annual costs of field sampling, lab analyses, and data interpretation for all six objectives. These values are annual costs in 1980 dollars. The range is from about 12,000 to over 2.5 million dollars per year.

The total costs over the six sampling periods, which range from 1 to 25 years, have been computed and are displayed in Table 11. The present worth of the costs as shown are the amounts of money that would have to be available in 1980 to cover the future costs escalating over time at the interest rates given for field sampling, lab analysis and data interpretation costs. The total values are seen to vary from $11,158 for a one-year design objective to $9,336,343 for a five-year research program covering 3 cities.

MODIFIED PLAN FOR OBJECTIVE I

On the suggestion of ERG, the objective I program was modified to include sampling fewer storms and interevent dry period sorties per year but more constituents. It was ERG's intuitive feeling that much more scientific information could be had under that program; and reductions in total cost should result, since added lab analyses costs would be less than the foregone, more expensive, field sampling costs.

Accordingly, let us see the effects on sample numbers and costs of

TABLE 7

SUMMARY OF FIELD AND LAB PROGRAMS FOR THE SIX OBJECTIVES

Objective*	Sorties/ year	Samples/ yr, n_a	Duration, yr, t	Total Samples, n	No. of Constituents, n_c	Lab Tests/ yr, n_y	Total Lab Tests, n_t
I	675	10,800	5	54,000	10	108,000	540,000
II	42	174	1	174	5	870	870
III	70	190	1	190	3	570	570
IV	4	24	1	24	5	120	120
V	200	600	25	15,000	3	1,800	45,000
VI	30	90	25	2,250	3	270	6,750

*Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

TABLE 8

LABORATORY ANALYSES ASSUMED AND LAB COSTS FOR EACH OBJECTIVE

Constituent	Lab Cost/Test, $	Objective* I	II	III	IV	V	VI
BOD	13.50	x	x	x	x	x	x
Coliforms	14.00	x	x			x	
Dissolved Oxygen	6.00	x			x		x
Nitrate-N	16.50	x	x		x		
Total organic-N	18.00	x					
pH	2.50	x			x		
Phosphate, dissolved ortho-	11.00	x					
Dissolved solids	6.50	x					
Suspended solids	6.00	x	x	x	x	x	x
Total solids	6.50	x	x	x			
Unit Cost per Sample, $**		100.50	56.50	26.00	44.50	33.50	25.50

*Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

**Unit costs derived from 1980 ERG price data; figures are approximate.

TABLE 9

INTERPRETIVE ANALYSES COSTS AND EFFORT FOR EACH OBJECTIVE

	Man-hours*			
Objective**	Senior Scientist	Research Scientist	Lab Personnel	Cost/yr, $***
I	675	2700	2700	238,275
II	42	84	168	12,012
III	70	140	280	20,020
IV	4	8	16	1,144
V	200	400	800	57,200
VI	30	60	120	8,580

*Assumptions: Senior Scientist Effort = 1/2 day/4 sorties
Research Scientist Effort = 1/2 day/sortie
Laboratory Staff Effort = 1 day/4 sorties for n_c less than 6 (Objs. II-VI)
1 day/2 sorties for n_c = 6 or more (Obj. I)

**Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

***Senior Scientist Rate = $53/hr
Research Scientist Rate = $41.50/hr
Lab Staff Rate = 33.50/hr
(1980 rate data derived from ERG price data; figures are approximate.)

TABLE 10

TOTAL ANNUAL COSTS FOR SAMPLING AND INTERPRETATION FOR EACH OBJECTIVE

	Partial Monitoring Costs, $/yr			
Objective*	Field Sampling**	Laboratory Analyses	Interpretation	Cost/yr, $
I	1,194,165	1,085,400	238,275	2,517,840
II	69,734	9,831	12,012	91,577
III	116,522	14,820	20,020	151,362
IV	5,852	5,340	1,144	12,336
V	404,600	20,100	57,200	481,900
VI	60,690	2,295	8,580	71,565

*Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

**ERG advises that automatic sampling equipment, while being 80 to 95 percent reliable, will nonetheless require 10 percent more effort (cost) for summertime maintenance and 20 percent more for wintertime maintenance. Inexperienced crews could result in maintenance effort nearly double that of experienced crews.

TABLE 11

TOTAL PERIOD COSTS FOR MONITORING TO SATISFY EACH OBJECTIVE

	Sampling	Present Worth of Sampling Costs, 1980$			
Objective*	Period, t, yrs	Field Sampling(9%)	Lab Analyses(12%)	Interpretation(12%)	TOTAL
I	5	4,564,791	3,912,624	858,928	9,336,343
II	1	63,976	8,778	10,725	83,479
III	1	106,901	13,232	17,875	138,008
IV	1	5,369	4,768	1,021	11,158
V	25	3,974,216	157,647	448,628	4,580,491**
VI	25	596,132	18,000	67,294	681,426***

*Objective I = Scientific research.
Objective II = Definition of problems.
Objective III = Analysis of alternative solutions.
Objective IV = Design of abatement measures.
Objective V = Regulatory compliance monitoring.
Objective VI = Operational performance monitoring.

**Costs in perpetuity (actually) = $5,139,722.
***Costs in perpetuity (actually) = $ 764,958.

halving the number of storms per year and dry period samples per event from 20 to 10, and adding constituents to both 14 and the full 20 as presented in Table 6.

Table 12 summarizes the modified objective I program. The dry period sorties have remained the same as before (135 per year) but the storm period sorties have been reduced from 540 per year to 270 per year. The dry period samples have been halved from 2,700 per year to 1,350 per year; and the storm period samples have been reduced from 8,100 to 4,050 per year.

However, while the field sampling and data interpretation requirements are lower, (because the numbers of sorties are lower), the laboratory analyses requirements will be greater (because the numbers of constituents are greater). The field, laboratory, and interpretation requirements for the modified program are given in Table 13.

Table 14 summarizes the effects of modifying the program on annual costs and total, 5-year program costs. Subprogram A is a program in which only 14 of the 20 constituents might be analyzed. Subprogram B includes all 20 constituents.

The total annual costs of the first objective I program was over 2.5 million dollars per year (Table 10). In the modified program, the annual costs for 14 constituents rather than 10, but from fewer events, would be 1.1 million dollars per year. For 20 constituents the annual cost would be 1.8 million dollars. The present worth values for the five-year period are the original 9.3 million dollars, versus 4.2 million dollars for subprogram A and 6.8 million dollars for subprogram B.

TABLE 12

MODIFIED OBJECTIVE I PROGRAM

Program Characteristics	Sorties Requires/Year	Samples Required/Year
n_u = 3 urban areas	Dry Period:	Dry Period:
n_w = 3 watersheds	$n_u n_w n_f n_b$ =	sorties x n_d =
n_f = 3 sites/w-shed	3x3x3x5	135x10 =
n_e = 10 storms/yr	= 135	1,350
n_s = 15 samples/storm	90 Type 1 + 45 Type 2	
n_b = 5 dry-per. events/yr	Storm Period:	Storm Period:
n_d = 10 dry-per. samples/ev.	$n_u n_w n_f n_e$ =	sorties x n_s =
t = 5 years	3x3x3x10	270x15 =
	= 270	4,050
	135 Type 3 + 135 Type 4	

TABLE 13

FIELD, LABORATORY, AND INTERPRETATION EFFORT FOR MODIFIED OBJECTIVE I

Field Sampling	Laboratory Analyses	Data Interpretation
8100 miles 23,040 man-hours 11.1 man-years @ $0.35/mile and @ $28/man-hour	Subprogram A 14 constituents in 4,050 wet samples/yr and in 1,250 dry samples/yr @ $46/wet sample and $92/dry sample	Senior Scientist 405 man-hours @ $53/m-hr
	Subprogram B 20 constituents in 4,050 wet samples/yr and in 1,350 dry samples/yr @ $156/wet sample and $312/dry sample	Research Scientist 1620 man-hours @ $41.50/m-hr
		Lab Personnel 1620 man-hours @ $33.50/m-hr

TABLE 14

ANNUAL COSTS AND OVERALL PRESENT WORTH FOR MODIFIED OBJECTIVE I PROGRAM

Annual costs, $/yr

Subprogram	Field Sampling	Lab Analyses	Interpretation	Total, $/yr
A (n_c = 14)	647,955	310,500	142,965	1,101,420
B (n_c = 20)	647,955	1,053,000	142,965	1,843,920

Present Worth, $1980

Subprogram	Field Sampling(9%)*	Lab Analyses(12%)	Interpretation(12%)	Total, 1980$
A (n_c = 14)	2,520,319	1,119,283	515,357	4,154,959
B (n_c = 20)	2,520,319	3,795,829	515,357	6,831,505

*Inflation rates in parentheses are taken from Figures 1 and 2.

VII. SUMMARY AND CONCLUSIONS

SUMMARY DISCUSSION

Storm and snowmelt runoff from urban lands contains quality constituents which under certain circumstances occur in damaging, polluting levels. These circumstances, the phenomena and processes making runoff quality levels what they are, remain so poorly understood and are so inherently random and variable that only field measurements can be relied upon to characterize their behavior.

Sometimes it is necessary or desirable to learn in great spatial and temporal detail what the processes are that cause deterioration in the quality of urban runoff. In these cases an expensive, comprehensive field monitoring program must be launched to learn the required information. By contrast there are other instances in which a planner or city engineer, for example, may wish to learn whether or not a particular constituent or several constituents in storm runoff attain potentially damaging levels in his or her community. In this case, a fairly cursory and modest monitoring program will suffice to satisfy this objective.

Statistical theory provides some guidance to the numbers of samples required for satisfying various objectives for monitoring. But a number of arbitrary variables, such as the degree of confidence or acceptable error must be specified by the monitoring program designer. Moreover, many statistical tests require that characteristics of some sampled data already be available, such as the sample standard deviation in the t-test, before they can be used to give guidance on the acceptability of sample numbers for still further monitoring.

Numbers of samples for example programs specified for satisfaction of six different monitoring objectives ranged from 24 for a design objective to 54,000 for a comprehensive scientific, process-monitoring objective. Costs for monitoring varied proportionately, from $11,158 to $9,336,343, for the two respective objectives.

Detailed sampling and laboratory analyses are expensive. Sometimes the expense of sampling and analysis will prove prohibitive. The major point of this work, however, is that curtailment of monitoring for reasons of expense or for any other reason can mean that the original monitoring objective cannot be met. If one sets out to elucidate runoff quality processes or to build and verify a model of these causative processes and finds that he can afford only $75,000 per year to do the indicated monitoring, then in fact he cannot adequately define processes or verify his model. The best he can hope for is to monitor operational performance, to gather information on what happened, what the effects were, and not what caused those effects to be realized. The information gained is likely not to be sufficient to allow prediction of what will happen in other places or under changed conditions.

Urban runoff quality in the early 1980's is not considered a high priority item of national environmental interest, except in those instances where urban runoff is conveyed and discharged through systems. Runoff quality control remains a new field of engineering endeavor. It is natural, therefore, that design criteria, discharge standards and receiving water criteria, and scientific knowledge about this field are still preliminary, incomplete, and rudimentary. For these same reasons, however, it is virtually axiomatic that mistakes may be made, monitoring and construction monies may occasionally be misspent on unnecessary or unsupported or inadequate facilities or monitoring programs. It is to be hoped that this guidebook will assist in keeping straight what is adequate monitoring for separate objectives and what is not.

CONCLUSIONS

1. Urban runoff may be sampled for a variety of reasons, i.e., to satisfy any of a number of objectives. The objective in each case should be stated explicity and restated as necessary in concrete, numerical terms so a program of sampling can be devised that will clearly satisfy that objective.

2. Sampling in the field is expensive. Laboratory analyses are both expensive and relatively inexpensive, depending on which constituents are being analyzed. But economies of scale in laboratory analyses are available, and opportunities should be sought for learning more for the same total budget by having more analyses performed and sampling fewer events in the field.

3. Statistical theory that exists today is somewhat useful in guiding decisions about sampling program design. Certain concepts from statistics, such as the existence of Type I and Type II errors, should be appreciated since they are axiomatic and unavoidable. Statistical tests such as the t-test, the F-test, and the chi-square test are useful and instructive and worthy of consideration, but they do not function usefully in the absence of any prior sample information. That is very frustrating but inescapable. Statistical theory on the whole is not developed sufficiently to assist meaningfully in multiple-site,

multiple-phenomena, multiple-constituent, stochastic-process sampling.

4. Mathematical models existing today for estimation of urban runoff quality loads do not have a solid scientific base. Some do not require or even give regard to the satisfaction of mass continuity. Directly as a result, they are either difficult or impossible to calibrate or they must be recalibrated with practically every use. Because they are so limited in transferability and even the most basic level of scientific rigor, they cannot be relied upon solely to replace or obviate the need for local quality sampling, regardless of the objective.

5. Research needs to be done to formulate the physics and chemistry of urban runoff behavior at several levels of sophistication and approximation. Just as it is not necessary to monitor each drop of rainfall, each clay particle, or each liter of runoff to design a detention basin, it is not necessary to model each of these things rigorously to design the basin without sample results. But design "theory" does need to be developed that is consistent with scientific principles, that allows for variability in behavior both chemically and physically, both quality-wise and hydraulically, and that approximates at least the continuity of mass and the processes of scour and deposition and chemical equilibria and rates of change of concentration with time and with changes in chemistry of the incoming and surrounding fluid. Current models in wide use are predicated on an assumed first-order decay or wash-off or first-flush phenomenon, which data from Denver and other places have clearly indicated does not always occur. Concentrations or masses of pollutants in runoff do _not_ always decrease exponentially or at all with time after the onset of rainfall. Current theory on the behavior of urban runoff and its quality constituents is poor at best and misleading or even wrong at worst. Hydraulic behavior is much better understood, and it is approximated adequately at various levels of study detail, i.e., for use with varying scientific, planning, design, or monitoring objectives. Quality theory should be developed and advanced in a commensurate fashion.

6. There is no guarantee that water quality data, no matter how carefully collected, will be transferable to other areas and other circumstances. However, because data collection and interpretation is so expensive and time-consuming, EPA is encouraged to maintain a data base, such as that now started at the University of Florida, wherein runoff quality data from across the nation may be stored

and accessed by subsequent users. But it is recommended that data so stored be characterized by the contributor as to their statistical properties (numbers of samples, variance, and the like) and as to the conditions and purposes under and for which they were collected.

APPENDICES

I. PERCENTILE VALUES FOR THE STUDENT'S t DISTRIBUTION

II. PERCENTILES OF THE F DISTRIBUTION

III. PERCENTILE VALUES FOR THE CHI-SQUARE DISTRIBUTION

Sources:

1. Haan, Charles T., Statistical Methods in Hydrology, The Iowa State University Press, Ames, 1977, 378p.
2. Lindgren, B.W., and G.W. McElrath, Introduction to Probability and Statistics, The Macmillan Company, New York, 1959, 277p.

APPENDIX I

PERCENTILE VALUES FOR THE STUDENT'S t DISTRIBUTION

Degrees of Freedom	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.60}$
1	63.7	31.8	12.7	6.31	3.08	1.38	1.00	.325
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.277
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.267
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.265
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.263
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.262
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.261
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.260
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.259
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.258
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.258
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.257
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.257
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.256
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.256
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.255
50	2.68	2.40	2.01	1.67	1.30	.849	.680	.254
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.254
80	2.65	2.38	2.00	1.67	1.30	.847	.678	.254
Infinite	2.58	2.33	1.96	1.64	1.28	.842	.674	.253

APPENDIX II

NINETIETH PERCENTILES OF THE F DISTRIBUTION

Denominator Degrees of Freedom	Numerator Degrees of Freedom 2	4	5	6	7	8	9	10	15	30	120
2	9.00	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.46	9.48
4	4.32	4.11	4.05	4.01	3.98	3.95	3.93	3.92	3.87	3.82	3.78
5	3.78	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.24	3.17	3.12
6	3.46	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.80	2.74
7	3.26	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.56	2.49
8	3.11	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.38	2.31
9	3.01	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.25	2.18
10	2.92	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.15	2.08
15	2.70	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.87	1.79
30	2.49	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.72	1.61	1.50
120	2.35	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.54	1.41	1.26

APPENDIX II (Continued)

NINETY-FIFTH PERCENTILES OF THE F DISTRIBUTION

Denominator Degrees of Freedom	Numerator Degrees of Freedom										
	2	4	5	6	7	8	9	10	15	30	120
2	19.0	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.5	19.5
4	6.94	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.75	5.66
5	5.79	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.50	4.40
6	5.14	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.81	3.70
7	4.74	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.38	3.27
8	4.46	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.08	2.97
9	4.26	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.86	2.75
10	4.10	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.84	2.70	2.58
15	3.68	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.25	2.11
30	3.32	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.84	1.68
120	3.07	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.55	1.35

APPENDIX II (Continued)

NINETY-SEVEN AND ONE-HALFTH PERCENTILES OF THE F DISTRIBUTION

Denominator Degrees of Freedom	Numerator Degrees of Freedom										
	2	4	5	6	7	8	9	10	15	30	120
2	39.0	39.2	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5
4	10.6	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.46	8.31
5	8.43	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.23	6.07
6	7.26	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.07	4.90
7	6.54	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.36	4.20
8	6.06	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	3.89	3.73
9	5.71	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.56	3.39
10	5.46	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.31	3.14
15	4.77	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.64	2.46
30	4.18	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.07	1.87
120	3.80	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.94	1.69	1.43

APPENDIX II (Continued)

NINETY-NINTH PERCENTILES OF THE F DISTRIBUTION

Denominator Degrees of Freedom	Numerator Degrees of Freedom										
	2	4	5	6	7	8	9	10	15	30	120
2	99.0	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.5	99.5
4	18.0	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.2	13.8	13.6
5	13.3	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.72	9.38	9.11
6	10.9	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.23	6.97
7	9.55	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	5.99	5.74
8	8.65	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.20	4.95
9	8.02	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.65	4.40
10	7.56	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.25	4.00
15	6.36	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.21	2.96
30	5.39	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.39	2.11
120	4.79	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.19	1.86	1.53

APPENDIX III

PERCENTILE VALUES FOR THE CHI-SQUARE DISTRIBUTION

Degrees of Freedom	$X_{.99}$	$X_{.975}$	$X_{.95}$	$X_{.90}$	$X_{.50}$	$X_{.10}$	$X_{.05}$	$X_{.025}$	$X_{.01}$
1	6.63	5.02	3.84	2.71	.455	.016	.0039	.0010	.0002
3	11.3	9.35	7.81	6.25	2.37	.584	.352	.216	.115
5	15.1	12.8	11.1	9.24	4.35	1.61	1.15	.831	.554
6	16.8	14.4	12.6	10.6	5.35	2.20	1.64	1.24	.872
7	18.5	16.0	14.1	12.0	6.35	2.83	2.17	1.69	1.24
8	20.1	17.5	15.5	13.4	7.34	3.49	2.73	2.18	1.65
9	21.7	19.0	16.9	14.7	8.34	4.17	3.33	2.70	2.09
10	23.2	20.5	18.3	16.0	9.34	4.87	3.94	3.25	2.56
12	26.2	23.3	21.0	18.5	11.3	6.30	5.23	4.40	3.57
14	29.1	26.1	23.7	21.1	13.3	7.79	6.57	5.63	4.66
16	32.0	28.8	26.3	23.5	15.3	9.31	7.96	6.91	5.81
18	34.8	31.5	28.9	26.0	17.3	10.9	9.39	8.23	7.01
20	37.6	34.2	31.4	28.4	19.3	12.4	10.9	9.59	8.26
25	44.3	40.6	37.7	34.4	24.3	16.5	14.6	13.1	11.5
30	50.9	47.0	43.8	40.3	29.3	20.6	18.5	16.8	15.0
40	63.7	59.3	55.8	51.8	39.3	29.1	26.5	24.4	22.0
50	76.2	71.4	67.5	63.2	49.3	37.7	34.8	32.4	29.7
60	88.4	83.3	79.1	74.4	59.3	46.5	43.2	40.5	37.5
70	100.4	95.0	90.5	85.5	69.3	55.3	51.7	48.8	45.4
80	112.3	106.0	101.9	96.6	79.3	64.3	60.4	57.2	53.5
100	135.8	129.6	124.3	118.5	99.3	82.4	77.9	74.2	70.1

APPENDIX IV

STATISTICAL GUIDANCE TO
SATISFACTION OF
SIX URBAN RUNOFF QUALITY MONITORING OBJECTIVES

Prepared by
Camp Dresser & McKee Inc.
Annandale, Virginia

Edited by
Michael B. Sonnen

Source: Dendrou, S.A., C.C. Clarkson, and L.A. Roesner, Statistical Aspects of Wet - Weather Monitoring, Presented to The M.B. Sonnen Co., Inc., CDM, Annandale, 25 October 1980, 58p.

OBJECTIVE 1: Sampling to Improve Fundamental Understanding of Urban Runoff Quality Sources, Determinants, Mechanisms of Change, and Constituent Interactions

This research objective is aimed at gathering new scientific knowledge. The aim of the objective is basically to help us gather data in a systematic way that will enable us to extend our knowledge and improve our fundamental understanding. To learn "truth" about any process or phenomenon first involves hypothesizing a relationship, process, or mechanism about the phenomenon that we wish to describe, i.e., the transformation of pollutant loads at rest on urban watersheds into urban runoff pollutant loads at "downstream" points. Furthermore, the researcher should attempt to describe this relationship analytically based on first principles. The degree of detail in which the process or mechanism is described or to which parameters that affect the process are to be represented spatially or temporally is a matter of choice for the investigator. The point is: The investigator must postulate, before measuring, what he expects to find in terms of some analytical description. The objective of the sampling then is to test the hypothesis that this analytical expression represents the phenomena that are occurring on the watershed during the runoff period.

Here the question is: How well does the postulated analytical "model" represent reality under all the conditions for which it was developed. Depending on the model at hand, the unknown of interest may be the flow-averaged concentration of various constituents, the mass rate of flow of a particular constituent during a storm, the total load per storm, per season, per year, some chemical equilibrium expression, or some other relationship. The population then is the infinite number of past and future realizations of this variate, x, and a sample set is any number of observations therefrom, x_{o_i}. The corresponding predicted or estimated values from the analytical expression are denoted by x_{e_i}.

This is a situation in which some goodness-of-fit test should be applied to determine whether the theoretical model is to be accepted or rejected at a given level of significance, a.

Example: Determine the smallest statistically acceptable sample size to test the representativeness of a predictive model of flow-weighted BOD concentrations at the 10 percent level of significance.

Additional assumptions are necessary concerning the nature of the acceptable discrepancy between model predictions and observations, and on the level of required accuracy of the prediction. The following assumptions are made:

1. The error is uniformly distributed over all observations, over the complete range of prediction of the model. This

means that the predictive model is assumed equally accurate within its predictive range. Otherwise, subranges should be tested separately. If a model is not supposed to predict first flush conditions, for example, data from first flush conditions should not be included.

2. The predictions should be within ±d mg/l of the observed values. This means that the expected standard deviation is s = d.

3. The largest acceptable average error = b mg/l.

These assumptions are translated in the following statistical statements. The error e_i, i.e., the difference between observed and predicted variates is normally distributed with mean zero and variance S^2.

$$e_i = x_{o_i} - x_{e_i} \qquad \text{(IV-1)}$$

Estimates of the mean and variance of the error are given by:

$$\bar{e} = \frac{\sum e_i}{n} \qquad \text{(IV-2)}$$

$$s^2 = \frac{\sum e_i^{\,2}}{n-1} \qquad \text{(IV-3)}$$

where n is the sample size. The allowable confidence interval from condition 3 above is b which can be equated to the confidence interval for $\bar{e}$ through Equation 11 in Chapter IV:

$$b = \frac{t_c s}{\sqrt{n-1}} \qquad \text{(IV-4)}$$

This assumes that the mean error of the population (as opposed to the sample) is zero. Rewriting Equation IV-4 and imposing condition 2, (s = d), we can describe a value that the Student t must display for a given number of degrees of freedom (v = n-1) to meet our accuracy criterion.

$$t_v = \frac{b}{d}\sqrt{v} = B\sqrt{v} \qquad \text{(IV-5)}$$

Finally, if we specify b as 1 mg/l and d as 4 mg/l, B = 0.25.

The value of t_v, from Equation IV-5, can be plotted as a fraction of the number of degrees of freedom. If the Student's t variate at a given level of significance, a, is also plotted, the intersection of the Student's t curve and the t_v curve from Equation IV-5 will intersect at the required number of samples. This is illustrated in Figure IV-1. The actual minimum sample size is a direct function of the level of significance and the factor B, which is the ratio of the acceptable average error, to the acceptable error at every measurement. For the particular value of B = 0.25, the minimum acceptable sample size is n - 1 = 46 hence n is about 50. This number is seen to be quite sensitive to the B factor.

One further test might be required, that is whether the standard deviation of the error is less than d. To test this, we could use Equation 14 of Chapter IV which can be written as:

$$\frac{S^2}{s^2} < \frac{(n-1)}{\chi^2(a/2,n-1)} \qquad \text{(IV-6)}$$

To develop the ratio $(n-1)/\chi^2$ as a function of n-1, we note from Figure IV-2, which shows plots of χ^2 vs n-1 for various values of a, that we can approximate χ^2 as a straight line, i.e.:

$$\chi^2(a,\ n-1) = k_a(n-1) \qquad \text{(IV-7)}$$

Substituting this relationship into Equation IV-6 gives:

$$\frac{S^2}{s^2} < \frac{1}{k_a} \qquad \text{(IV-8)}$$

a condition that is independent of sample size and dependent only on the level of confidence. Therefore, the sample size for testing our analytical expression can be viewed as being dependent only on our test of the mean error between the estimated (predicted) analytical values and the observed values from the field measurements (Equation IV-5 and Figure IV-1).

OBJECTIVE 2: Sampling to Define the Existence and Scope of Urban Runoff Quality Problems in Existing Urban and Urbanizing Areas

For the case in which it is desired to compare one type of area with another (one urban watershed with another urban watershed, or an urban

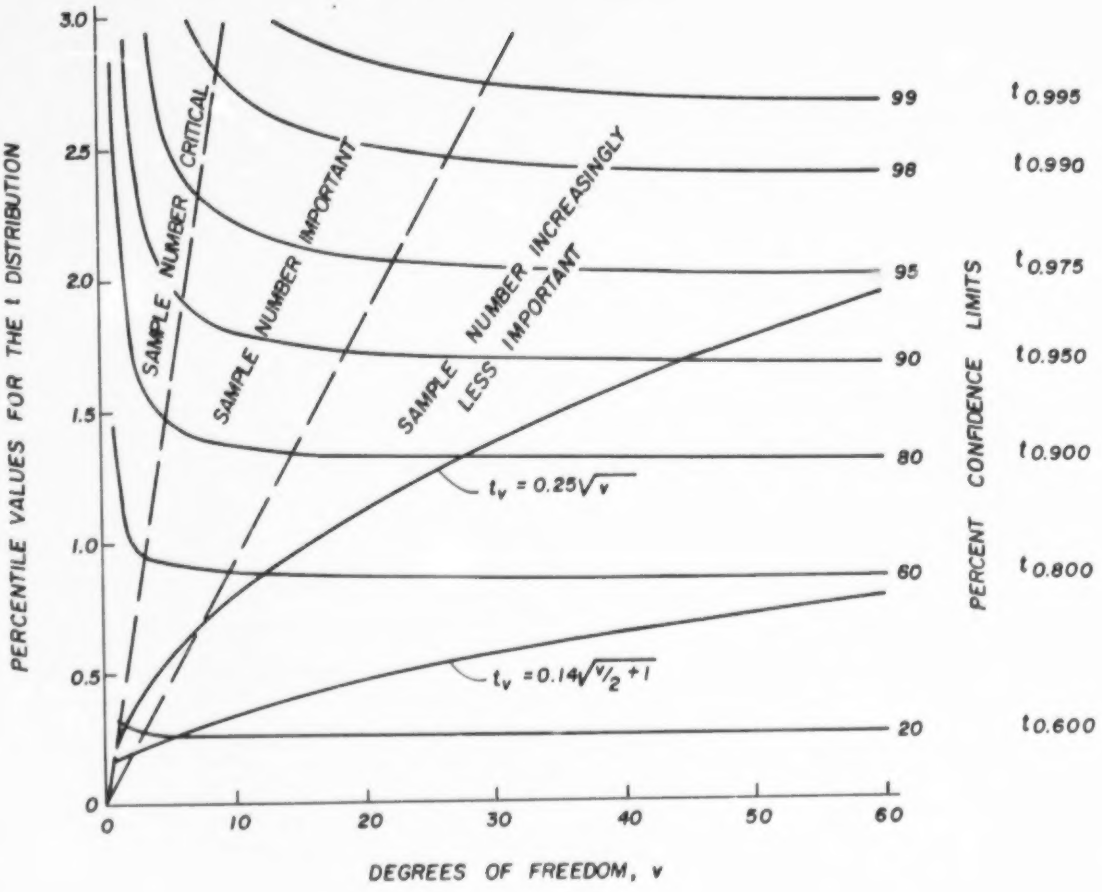

Figure IV-1. Example Student's t Distributions

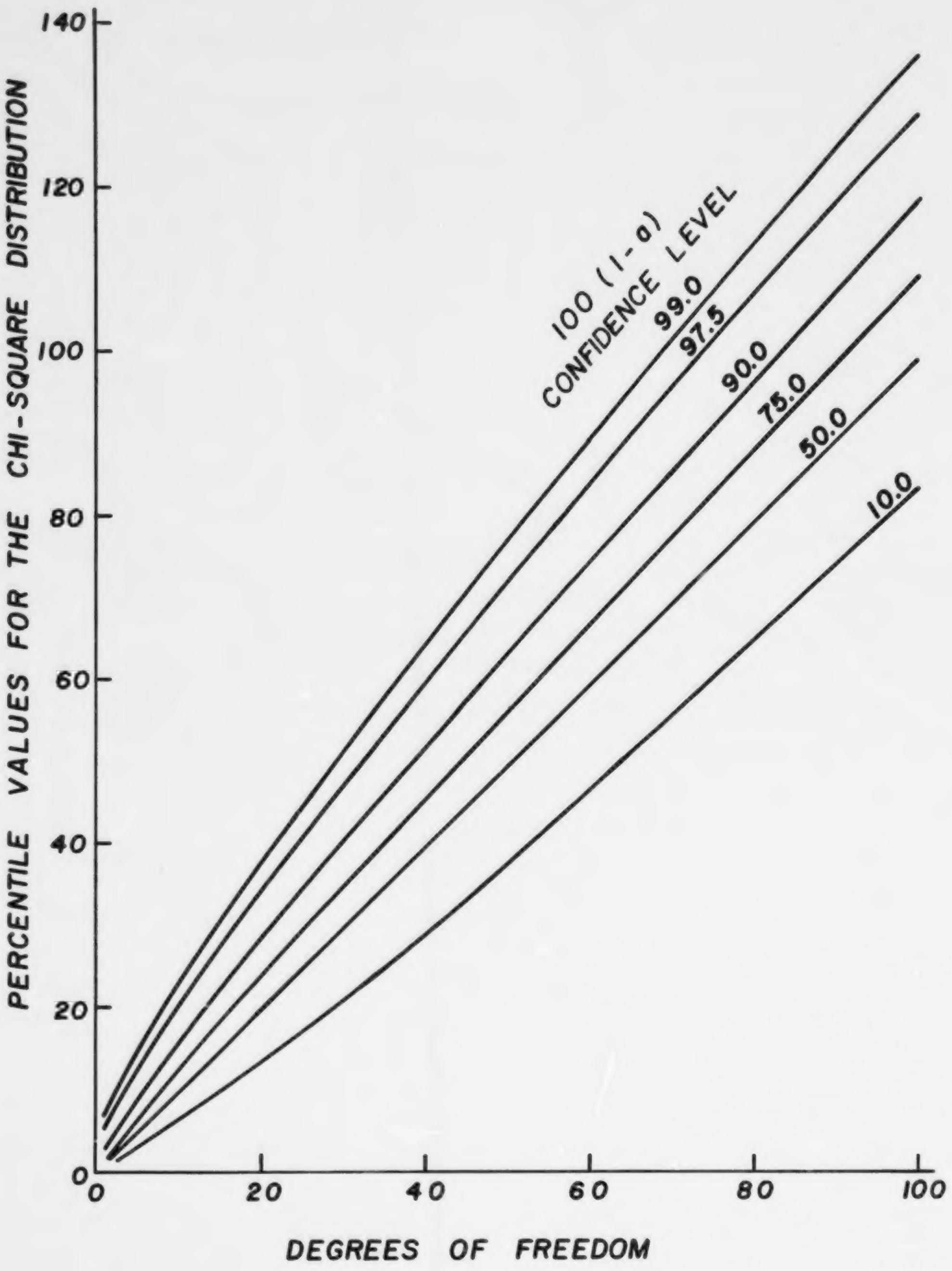

Figure IV-2. Example Chi-Square Distributions

watershed with an undeveloped watershed), one may again make tests to examine the behavior of one discharge relative to another during the same event or time period, or to examine relative behavior over a period of time.

This is the case where we are comparing two populations, namely the variates extracted from two watersheds or two discharges. We want to determine if there are signigicant differences in the respective levels of pollutants discharged from those separate sources. Provided that the two sample sets are each representative and comparable, a confidence interval test may be performed for the difference between the two population means. For example, provided that the two watersheds are from the same hydrologic and geomorphologic region, a comparable variate is the total load per event per acre, e.g., x_1 and x_2 lbs of PO_4/acre/event. Then, the 1-a confidence interval for the difference of population means is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1 - a/2;\ n_1 + n_2 - 2)} \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \text{(IV-9)}$$

where S_p^2 is the weighted variance:

$$S_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} \qquad \text{(IV-10)}$$

Assuming that the same number of samples will be taken from both watersheds, $n_1 = n_2 = n$, the test against no difference gives:

$$|t| > t_{(1 - a/2,\ 2n - 2)}$$

$$|t| = |\bar{x}_1 - \bar{x}_2| \sqrt{\frac{n}{s_1^2 + s_2^2}}$$

$$|\bar{x}_1 - \bar{x}_2| \sqrt{\frac{n}{s_1^2 + s_2^2}} > t_{(1 - a/2,\ 2n - 2)} \qquad \text{(IV-11)}$$

which is of the same form as for the example of Objective No. 1, provided estimates can be produced for the acceptable average difference $\bar{x}_1 - \bar{x}_2$, and the corresponding standard deviations s_1 and s_2.

As an example, suppose we are monitoring runoff from an agricultural and a light industrial area, and we want to compare the mean concentrations of soluble phosphorus from the two watersheds. Suppose further, that the literature indicates the following statistics:

Agriculture: $\bar{x} = 0.3$ mg/l, $s_1 = 0.05$ mg/l

Light Industry: $\bar{x}_2 = 0.2$ mg/l, $s_2 = 0.5$ mg/l

Rewriting Equation IV-11:

$$\frac{t_{(1 - a/2,\ 2n - 2)}}{\sqrt{n}} \leq \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{s_1^2 + s_2^2}} \qquad \text{(IV-12)}$$

Now let us denote t_v as the value that the computed Student t variate would take for a given ratio $\left|\bar{x}_1-\bar{x}_2\right| / \sqrt{s_1^2+s_2^2}$ and v degrees of freedom ($v = 2n - 2$).

Then:

$$t_v = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{s_1^2 + s_2^2}} \sqrt{v/2 + 1} \qquad \text{(IV-13)}$$

Substituting the expected data from our example problem defines t_v as a function of v only, i.e.:

$$t_v = \frac{0.1}{0.7} \sqrt{v/2 + 1} = 0.14 \sqrt{v/2 + 1} \qquad \text{(IV-14)}$$

The degrees of freedom v at which t_v, as defined by Equation IV-14, equals t (1 - a/2, v), as defined by the Student t distribution at significance level a, will define the minimum sample size required to test the means.

In Figure IV-1, we showed the Student's t plot under Objective 1 with Equation IV-14 added as well. The figure indicates that for a confidence limit of 60 percent, v would have to be about 65. Since $v = 2n-2$, this means the minimum sample size would be 34. At 90 percent confidence, the Student t curve and t_v curve will intersect at $v = 551$ or $n = 276$ observations. We note that the sample size is very sensitive to the ratio of $\left|\bar{x}_1-\bar{x}_2\right|$ to $\sqrt{s_1^2 + s_2^2}$. For example, if $\left|\bar{x}_1-\bar{x}_2\right|$ is 0.3 mg/l rather than 0.1 mg/l as in our example, the required sample size would be reduced to about 10 observations instead of 276. Therefore, external considerations regarding hydrology, geomorphology, and other things may be more important (constraining) than the strict statistical requirements.

OBJECTIVE 3: Sampling to Examine Alternative Runoff Quality Control Measures

This sampling objective could pertain to monitoring conducted as part of an analysis of treatment alternatives and preliminary design of a selected alternative. Sampling is done to support the adoption of the selected alternative. Here we might be asking whether street sweeping is a more effective control measure than detention ponds or swale drainage.

There are two needs. One is to compare runoff quality data with and without the control measure. A second is to compare performance data for one control measure with that of another control measure. Sampling variates that might be collected to evaluate performance are: load per storm, monthly or seasonal loads, mean concentration or maximum concentration per event.

There are two ways to gather data for this determination. One is to compare runoff data from a single watershed before and after a control measure has been implemented or installed, or to compare data from two similar watersheds, one with the control measure and one without. The advantage in the first case is that the same watershed is used but the storms are different. In the second case, the areas are not identical but the storms are the same. Moreover, less calendar time is required for sampling since the samples in the second case are taken in parallel rather than in series, and comparisons may be made as events occur. For end-of-pipe control measures, data can be collected at the influent and effluent for comparison. Whatever the method of sampling, the statistical question here is one of hypothesis testing; the hypotheses being that the means of the samples are the same and that the variances of the samples are the same.

For comparing removal amounts or removal efficiencies, one can test to estimate the difference in the mean concentrations of the influent and effluent or the mean value of the influent load and the effluent load on a monthly, seasonal, or annual basis. The determination of the difference in these means can be made at the corresponding level of significance attached to the estimate. Determination of the appropriate sample size follows the same procedure as the example in Objective 2. However, if the treatment measure is at all effective, the differences in the sample means (with vs without treatment) should be substantial so that the ratio of $|\bar{x}_1-\bar{x}_2|$ to $\sqrt{s_1^2 + s_2^2}$ is not so small (Equation IV-12); therefore, the sample size for comparison of alternatives is smaller. For example, suppose that some treatment measure applied to the industrial area discussed under Objective 2 was expected to reduce the mean soluble phosphorus concentration in the runoff from 0.2 mg/l to 0.001 mg/l with a standard deviation of 0.001 mg/l, the ratio in Equation IV-14 would be increased from 0.14 to about 0.4 (0.199/0.499) and the minimum sample size at the 90 percent confidence level would be about 35, as opposed to 276 in Objective No. 2.

The F-test might be particularly useful in investigating whether some treatment measure (e.g., detention basins or swirl concentrators) has had a significant effect in reducing the variability of pollutant concentrations. For example, if runoff from two watersheds shown over time to be equal in concentration variance via the F-test, are no longer equal after some treatment measure has been instituted in one watershed, i.e., concentrations show a smaller scatter around acceptable values, then the treatment may be considered to be effective.

One may want to test the hypothesis that the population variances, with or without the device are the same. The "null" hypothesis is: $S_1^2 = S_2^2$ and the test statistic is $F = s_1^2/s_2^2$ with the critical region defined at the a-level of significance by:

$$F_{(a,\ n_1,\ n_2)} < F < F_{(1-a,\ n_1,\ n_2)} \qquad \text{(IV-15)}$$

Taking $n_1 = n_2 = n$, $a = .025$ and considering that for a noticeable change in variability we should have at least $s_1/s_2 = 1.40$, thus we would have:

$$F = \frac{s_1^2}{s_2^2} = 1.40^2 = 1.96 \qquad \text{(IV-16)}$$

for which, Appendix II shows, n is about 35 ($F_{0.975,30,30} = 2.07$).

This would mean that we would have to monitor 35 storms from each watershed and consider as acceptable a 40 percent change in the variation of the particular variate. Again, the sample size is very sensitive to the acceptable value of s_1/s_2 and the level of significance.

OBJECTIVE 4: Sampling to Support the Final Design and Installation of a Chosen Abatement Measure

The purpose of collecting data in support of this objective can be viewed as calibration of the design algorithms used for "sizing" an abatement measure. All installations are sized on the basis of some design algorithm which may include variables that are dependent on local geography, hydrology, or land use. The statistical tool to be used here is a test of goodness-of-fit of calibration measurements to the design algorithm.

In the calibration procedure, model parameters that affect the quality of effluent from the abatement control measure can be varied until the investigator is satisfied, at a given level of significance, that the calibration is complete. One useful procedure would be to operate the design algorithm, compute behavior values, test them against observed values from the field with the test on the residual error described in Objective 1, and adjust the model parameters until a satisfactory calibration had been achieved. The number of samples taken to perform the goodness-of-fit test and to verify that the theoretical approximation is a good representation of the observed phenomena at a given level of significance, follows the procedure described under Objective 1.

For example, if we know that our design algorithm can simulate suspended solids reductions within ±10 mg/l on the average with a standard deviation of 15 mg/l, the ratio b/d in Equation IV-5 is 10/15 = 0.67. To calibrate at the 90 percent confidence level requires $n = v+1$ observations where v is determined as the point at which t_v and $t_{(.90,v)}$ intersect. This value is $v = 8$ or $n = 9$ observations.

OBJECTIVE 5: Sampling to Monitor Compliance with Regulatory Requirements

Two decision-making problems can be identified which pertain to a choice among alternatives. One is to decide whether there is compliance or violation. This is a discrete (yes-no, or 0-1) type of choice. The other is to quantify the extent to which there is compliance or violation.

There are two types of decision-making criteria, according to existing regulatory requirements. These are 1) the absolute, non-violatable threshold limit type and 2) the frequency and duration-of-exceedance type. It should be noted that the threshold type of criterion lends itself to the first, discrete, type of choice among alternatives, and it is thus simpler to use from an administrative enforcement viewpoint. Because the exceedance type is gaining favor and is so much more applicable to the highly variable urban runoff situation, only an example of that type is given.

Example: Monitor receiving waters (streams) with a frequency-type standard in effect: The coliform group of bacteria is not to exceed 70 per 100 ml as a monthly average value, nor to exceed this number in 20 percent of the samples examined during any month nor to exceed 2,400 per 100 ml on any day.

For wet-weather monitoring purposes, monthly frequencies are tied to the frequency of occurrence of storms and to the finite size of the population of storms during a given month. A monthly average value is interpolated here as meaning the true average of any given month's observations rather than the per month average of a yearly variation.

Assuming that n storms are monitored during a month, and that the correct per storm averages are determined, then the monthly averages $\bar{x}$ are obtained as the average of the per event averages x_i; and the standard states that $\bar{x} < 70$ counts/100 ml with a certain level of significance, a = 20 percent. This falls into the t-statistic category, and the procedure follows that of Objective 1.

$$t_{(1-a/2,\ n-1)} < \sqrt{n}\ \frac{(\bar{x} - 70)}{s} = \frac{b}{d}\sqrt{n} \qquad \text{(IV-17)}$$

If the maximum acceptable violation (b) is five counts/100 ml, and the tolerable standard deviation (d) is 10 counts/100 ml, then $t_{(1-a/2,n-1)} < 0.5\sqrt{n}$, which gives, from Figure IV-1, n=13 storms per month. (Note that is a large number of events for a single month. Several years of sampling may be necessary to obtain that many samples for each month of the year.)

OBJECTIVE 6: Sampling to Monitor Operational Performance or Effectiveness of Constructed Works or Management Practices

This objective is directly analogous to Objective 4, the support of final design. However, in the case of sampling for final design the objective was really to calibrate a design algorithm. Under this objective the purpose of sampling could be viewed as verifying the design algorithm. One method for doing this might be to use the design to predict performance under an actual design event and to compare these predictions with measurements taken during that event. This is a fairly straightforward application of the chi-square test for goodness-of-fit. The actual statistic that we will want to monitor will depend on the attribute used in the design algorithm. For example, the performance of a treatment process that is designed to reduce concentrations during a storm event can be compared with predicted concentrations from the design algorithm. In comparing a detention pond or street sweeping, on the other hand, the parameter of interest would probably be total load per storm. In that case we would want to composite samples over the storm to get a total load.

The example presented under Objective 4 is directly applicable to this case and indicates that nine observations are required to verify that this facility or management practice is operating as planned, at the 10 percent acceptance level or with 90 percent confidence.